UNIVERSITY OF NAIROBI.

GENETIC DIVERSITY AND PHYLOGENETIC ANALYSIS OF *Adansonia Spp.* IN THE KENYA/TANZANIA TRANSECT.

BY CATHERINE WAMBUI KAMAU

I56/71711/2014

Center for Biotechnology and Bioinformatics

College of Biological and Physical sciences

The University of Nairobi

A thesis submitted in partial fulfillment for the degree of Master of Science in Biotechnology in the Center for Biotechnology and Bioinformatics (CEBIB)

University of Nairobi.

November 2016

DEDICATION

To my family and dear friend, Louis Njora for their support and encouragement.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

- °C   Degrees Celsius
- µl   microliters
- AIC   Akaike Information Criterion
- AMOVA   Analysis of Molecular Variance
- Approx.   Approximately
- Bp   base pair
- DNA   Deoxyribonucleic acid
- EST   Expressed Sequence Tag
- H   Diversity
- Hrs   hours
- ID   Identification
- Inc.   Incorporation
- ITS   Internal Transcribed Spacers
- m   meters
- min   minutes
- ng   nanogram
- nrDNA   Nuclear Ribosomal DNA
- PCR   Polymerase Chain Reaction
- PCR-RLFP   Polymerase Chain Reaction-Restriction Fragment Length Polymorphism
- pmol   picomole
- QI   Quality Index
- rRNA   Ribosomal ribonucleic acid
- s   seconds
- sect   Section
- spp   species
- Tm   annealing temperature

ABSTRACT

Baobab is a widely used indigenous tree species in Africa. It is valued in the food and cosmetic industries for its high vitamin C content in its fruit pulp and its seed oil respectively. It has a high potential for domestication. However, baobab populations are declining due to climatic changes and human development activities. It is therefore necessary to carry out molecular characterization in order to understand the genetic variation and population clustering which will aid in developing conservation strategies for the species. This study evaluated the geographical distribution of *A. kilima* and *A. digitata* in the Kenya /Tanzania transect given the difficulty in distinguishing the species visually. It also determined the level of genetic variation and differentiation of 8 populations of *Adansonia* species sampled from the Kenya/Tanzania transect. The genetic variation was determined using 16 microsatellite markers at 16 loci in 62 trees. An analysis of Molecular Variance (AMOVA) showed a low level of genetic variation among populations: 96% within populations and 4% among populations. This was due to high gene flow among the populations. The total genetic differentiation coefficient was (PhiPT= 0.042, P < 0.001). A UPGMA dendogram and neighbor joining radial tree generated by DARwin 6 grouped the genotypes into 3 distinct clusters. Cluster 1 consisted of Kenyan populations, Cluster 2 both Kenyan and Tanzanian populations and Cluster 3 consisted of Tanzanian populations. There was some general correlation between clustering and geographic origin indicating some degree of genetic structuring in the populations. The average heterozygosity (H) values ranged from 0.1529 to 0.2736. The percentage of polymorphic loci was based on a 99% criterion and it ranged from 34.0299 to 63.8806. Kondo population was the most diverse with a mean diversity of 0.194 and 17 private alleles. The two species (*A. digitata and A. kilima*) being superficially similar, were differentiated on the basis of phylogenetic analysis. The nuclear ITS region (ITS-1, ITS-2), amplified using 2 primers (ITS LEU/4 and ITS S2F/S3R) was used to examine phylogenetic relationships within *Adansonia* and determine whether *A. digitata* evolved from *A. kilima*. Phylogenetic analysis of the ITS region was not informative enough to distinguish between *Adansonia digitata* and *Adansonia kilima* or determine whether *A. digitata* evolved from *A. kilima*.

CHAPTER ONE
INTRODUCTION

1.1 Background

Baobabs (*Adansonia spp.*) are significant iconic trees. They are known for their strange forms, immense size, source of food and as the subjects of myths and mysteries (Cruywagen et al., 2010). Their iconic shape, extraordinary longevity, ethnobotanical significance, economic importance and intriguing natural history have captured and resulted in popular imagination for many centuries (Pettigrew et al., 2012). Charles Darwin, in 1832, documented having observed baobab trees on St. Jago in the Cape Verde Islands (Pettigrew et al., 2012). He commented on their size and longevity. David Livingstone, during his account of his travels in Southern Africa, mentioned baobab trees several times in the 1840s and 1850s (Livingstone, 1861).

Baobab (*Adansonia digitata* L) is among the most widely used indigenous tree in sub-Saharan Africa (Munthali et al., 2012). It belongs to the Bombacaceae family which includes 6 tribes, about 30 genera and 250 species. Baobab belongs to the genus *Adansonia.* It has been found to have a wide geographical distribution; six species are endemic to Madagascar, one in Australia and two in Africa (Baum et al., 1998). The African baobab (*Adansonia digitata*) is a significant iconic tree with considerable ethnobotanical significance. Unlike the other eight species which are diploid, *Adansonia digitata* is tetraploid. *A. digitata* should therefore have a diploid ancestor however, no diploid species had been recorded in the African mainland until recently (Pettigrew et al., 2012). By examining variation in floral and pollen characteristics and chromosome number in specimens from different parts of Africa, Pettigrew and coworkers identified a new diploid species, *Adansonia kilima.* The new species was found to co-exist with *A. digitata* in Africa.

The African baobab occurs naturally in traditional agroforestry systems and is associated with the savannah, especially the drier parts (Assogbadjo et al., 2006). Different chromosome numbers for Baobab have been published but Baum & Oginuma, (1994) hypothesized that the African Baobab is an autotetraploid that originated from aneuploid reduction from 4x =176. However they recommended that additional research be done in order to clarify the cytology of Bombacaceae. The genus *Adansonia* was argued by Baum et al., (1998) to have originated in Madagascar and migrated to Africa before the breaking of West Gondwana blocks at the beginning of the Cretaceous by long distance dispersal. Within the species, the existence of a number of local types has been indicated. These types differ in vigor, size, habit, vitamin quality of the leaves and the quality of the fruits (Gebauer et al., 2002; Sidibe & Williams, 2002).

1.2 Problem Statement

The declining population of Baobab trees is quickly becoming a matter of great concern. The Baobab tree species are at risk of losing majority of their available habitats due to human development activities and climate change. The loss of seed dispersers could also threaten the survival of some baobab species. Large animals, especially elephants, greatly contribute to this dispersal. They eat the fruits, disperse the seeds and break the seed dormancy which encourages regeneration. Elephants however are at risk of extinction because of illegal poaching and habitat destruction and this could affect regeneration of Baobab and ultimately the populations of Baobab trees. There is also a lack of awareness among the local people on the importance of planting, protecting and managing under-utilized fruit species. There is therefore a need to carry out molecular characterization in order to determine the level of genetic diversity of the Baobab trees and develop appropriate conservation strategies for the species.

1.3 Justification of the Study

The Baobab is an example of an Indigenous wild fruit tree (IFT) and is of great importance in arid and semi-arid Africa, especially in regions where other fruit species cannot be cultivated easily. These trees contribute to nutrition, food security, health and provide a source of income for the local people. This is mostly during pre-harvest and food shortage periods where they serve as emergency food. In order to improve rural livelihoods, household incomes, economic opportunities and promote biodiversity conservation and sustainable use of resources in the tropics, domestication and commercialization of baobabs should be considered (Gebauer and Luedeling., 2013). The baobab tree is also a prime candidate for domestication in the semi-arid regions of Africa. Its fruits are economically important both locally and internationally, where they can be used as food, medicine and in pharmaceutical and cosmetic industries.

Studying the genetic diversity within a species is important in understanding how the species would respond to environmental changes. Genetically diverse trees are more likely to withstand adverse changes and this is important in ensuring survival and evolution of the species. It also provides information on the history and present population structure of the species. This knowledge can ultimately be utilized to design breeding, management and conservation strategies. Pettigrew et al., (2012) in his study did not clearly describe the geographical distribution of *A. kilima* in the East African region. Reproduction between the diploid (*A. kilima*) and tetraploid (*A. digitata*) trees may result in triploid offspring which often leads to infertility. This infertility may be due to the production of unviable, unbalanced, and semi-sterile gametes caused by a lack of homologous pairing. This leads to differences in fruit production among the trees which affects the communities who greatly utilize baobab fruits as a food source. There is therefore a need to demarcate the specific regions of coexistence between *A. digitata* and *A. kilima* in the Kenya/Tanzania transect.

1.4 Research questions

1. What is the level of genetic variation and polymorphism among and within the *Adansonia* populations?
2. Is there a partitioning of genetic diversity in *A. digitata* and *A. kilima*?
3. Can phylogenetic analysis of *Adansonia* species be used to confirm presence of *A. kilima* and its coexistence with *A. digitata* along the Kenya/Tanzania transect?

1.5 Hypotheses

1. There is low genetic polymorphism among and within *A. digitata* and *A. kilima* provenances sampled.
2. There is no partitioning of genetic diversity in *A. digitata* and *A. kilima*.
3. There is no presence of *A. kilima* in the Kenya/Tanzania transect.

1.6 Study objectives

1.6.1 Main objective

To define the spatial genetic structuring, distribution and variation of diploid *Adansonia kilima*, and tetraploid *Adansonia digitata* species along the Kenya/Tanzania transect.

1.6.2 Specific objectives

1. To determine whether genetic patterns and molecular analysis of the internal transcribed spacer region (ITS) reveals occurrence of *Adansonia kilima* along the Kenya/Tanzania transect.
2. To determine the level of genetic variation among the *Adansonia spp.* along the Kenya/Tanzania transect.
3. To determine the level of polymorphism of the *Adansonia* populations.

CHAPTER TWO
LITERATURE REVIEW

2.1 Origin and distribution of baobab

Baobab trees are iconic plants and they represent some of the most recognizable trees in the world (Cruywagen et al., 2010). They are recognized as Africa's 'upside-down tree' with a distinctive and bizarre appearance making them important landmarks throughout many African savannah landscapes (Gebauer & Luedeling, 2013). Baobab is the common name given to the nine species of the genus *Adansonia.* It belongs to the Bombacaceae family (Assogbadjo et al., 2006)**.** This generic name honors Michel Adanson, a French explorer and naturalist who described *A. digitata* in 1761 (Pettigrew et al., 2012).

Baobab trees have been found to live for more than 1000 years (Adrian et al., 2007). The *Adansonia* genus is characterized by massive trunks that are often bottle-shaped, palmate compound leaves, and a large, dry indehiscent fruit containing reniform seeds embedded in an edible pulp (Baum, 1995b). Ecologically, the baobab tree can withstand drought, has no significant diseases or pests and it does not require water or fertilizer for cultivation (Sanchez, 2011).

The genus *Adansonia* consists of 9 known species (Baum et al., 1998; Pettigrew et al., 2012). *A. gregorii* is endemic to the Kimberly region of North-Western Australia. *A. digitata* is native to mainland Africa and is closely related to *A. kilima* which was identified and described by Pettigrew et al (2012). *A. digitata* is known to have migrated throughout the tropics and to surrounding islands including Madagascar, by natural and human-mediated dispersal both terrestrially and overseas (Tsy et al., 2013). It is the oldest known tropical angiosperm species (Ndoye et al., 2013). The other 6 species: *A. perrieri, A. za, A. rubrostipa, A. grandidieri, A. suarezensis* and *A. madagascariensis* are endemic to Madagascar and they constitute a monophyletic radiation. The current distribution of

baobabs has been explained either by long-distance dispersal from Africa/Madagascar to Australia or in terms of vicariance during the breakup of the Gondwana (Baum et al., 1998).

2.2 Floral morphology and pollination of baobab

Adansonia species display a striking diversity in floral morphology which is used as a basis for the recognition of 3 distinct sections. *Adansonia sect. Adansonia* comprises of *A. digitata.* It has globose flower buds hanging down on long flower stalks that comprise both the pedicel and peduncle. *A. digitata* is known to be bat pollinated (Harris, 1959; Start, 1972) and is also visited by bushbabies (Baum et al., 1998). *Adansonia sect. Brevitubae* is composed of two Malagasy species. They include: *A. grandidieri* and *A. suarezensis.* They have ovoid flower buds borne on short erect stalks. During the dry season, their flowers are small, white and open. They are mainly pollinated by nocturnal mammals primarily fruit bats and lemurs (Tsy et al., 2013). *Adansonia sect. Longitubae* consists of the remaining 4 Malagasy species, *A. perrieri, A. za, A. rubrostipa, A. madagascariensis* and the Australian species, *A. gregorii.* They have long cylindrical flower buds that are borne either erect or horizontal. They are almost exclusively pollinated by large hawk moths (Baum et al., 1998). The Malagasy species are distinguished by bright-colored flowers (orange, red or yellow), which open during the wet season. They are also pollinated by hawk moths (Tsy et al., 2013).

2.3 Common uses of *Adansonia* species

The baobab is a multi-purpose tree species with a variety of uses. It supplies raw materials for hunting and fishing equipment and making musical instruments. Hollow baobabs historically have been used as prisons, burial sites, watchtowers, stables, storage rooms, shelters and for storage of water (Pettigrew et al., 2012). Baobab trees were known to ancient Egyptians and they were mentioned in travel accounts with relative frequency since the 1300s (Baum, 1995a).

Different plant parts are used as medicines, foods (Sidibe and Williams, 2002) and the bark fibers are used for rope-making, weaving and making of artistic materials (Gebauer et al., 2002). Every part of the tree has been reported to be useful especially in generating income for rural communities. The pulp of its fruit has very high vitamin content, almost ten times that of oranges (Gebauer et al., 2002). It is used as an appetizer, in seasoning foods and also to make juices. Baobab seeds contain adequate quantities of digestible carbohydrates and oil, crude proteins and high levels of lysine, thiamine, iron and calcium (De Caluwé et al., 2009).

Baobab trees are also used as cattle feed. This is important in the savanna areas especially in arid zones, where animals obtain most of their feed in the form of leaves and pods. In folk medicine, baobab pulp is used to treat fevers and dysentery. The pulp extract may also be applied as eye-drops in cases of measles. The leaves also form a component of many herbal remedies. A mash prepared from the dried powdered roots is used as a tonic in malaria patients. Sores may also be treated by the semi-fluid gum obtained from the baobab bark (Gebauer et al., 2002). The baobab tree has been identified as one of the most important edible savannah trees to be domesticated, conserved and valorized in Africa, due to its economic relevance at local to international scales (Gebauer and Osman, 2004; Jama et al., 2008; Raebild et al., 2011). It is also valued in the cosmetic industry for its seed oil (Munthali et al., 2012).

2.4 The African species of *Adansonia*

The African species of *A. digitata*, unlike the other 8 species that are diploid, is tetraploid and this creates a problem in regards to *Adansonia* origins (Baum & Oginuma, 1994). It was possibly reduced from an aneuploid chromosomic type. A number of predictions have been established to try and explain the origin of *A. digitata*. One possibility is that *Adansonia* could have centrally originated from Africa following indirect or direct

dispersal of the progenitor from the New World (Baum et al., 2004). Ancestors living in Africa could have been diploid with evolution of tetraploidy occurring after the divergence of Australian and Malagasy lineages. Alternatively, the diploid progenitor may have had its origins outside of Africa for example in Madagascar, with the tetraploid ploidy level evolving during or after dispersal to Africa. This hypothesis is shown in Figure 1 (Pettigrew et al., 2012).

However, after many decades of controversy, a phylogeographical analysis using Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RLFP) of DNA chloroplast fragments revealed that *A. digitata* probably originated from West Africa and migrated throughout the rest of the world (Ndoye et al., 2013). There is therefore a possibility that a diploid progenitor of the tetraploid is still existent in Africa.
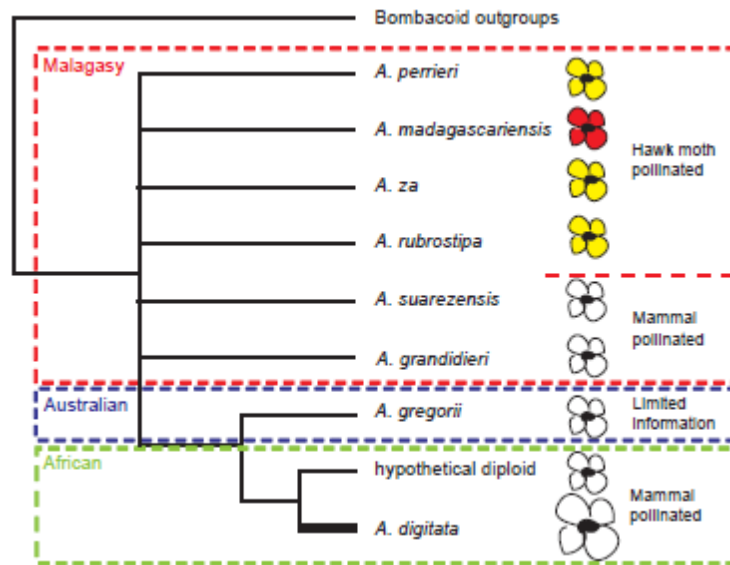
Figure 1: A hypothetical phylogeny that guided the search of an African diploid baobab. The thicker line represents the tetraploid lineage *(Pettigrew et al., 2012).*

2.4.1 The diploid progenitor of *A. digitata*

The hypothetical phylogeny illustrated in Figure 1 guided the search for an African diploid baobab species. All extant baobab species are diploid apart from *A. digitata*. As it is in most plants, tetraploidy is derived from the evolution of a diploid progenitor. There is no evidence of a tetraploid baobab species in Australia or Madagascar. An African diploid progenitor could therefore be hypothesized to have evolved into the tetraploid African species, *A. digitata*. In support of this hypothesis, phylogenetic analysis by (Pettigrew et al., 2012) identified a new diploid species, *A. kilima* which coexists with *A. digitata* in South-Eastern and Southern Africa (Gebauer & Luedeling, 2013). *A. digitata* and *A. kilima* were found to be morphologically similar. This suggested that tetraploidy evolved recently (Ndoye et al., 2013). The new species was identified through molecular analysis, chromosome analysis and examination of variation in floral and pollen characters.

This study described *A. kilima* as being superficially similar to *A. digitata.* However, the two can be differentiated on the basis of pollen, floral morphology and chromosome number. *A. kilima* is also restricted to moderate elevations (650-1500m) unlike *A. digitata* which prefers elevations below 800m (Pettigrew et al., 2012). The study by Pettigrew et al., (2012) also found *A. kilima* in the eastern slopes of Mt. Kilimanjaro, West to Northern Namibia, South to Southern Tanzania and in the Venda of South Africa. The current study seeks to evaluate the geographical distribution of *A. kilima* in the Kenya /Tanzania transect and clearly define the overlap between the diploid (*A. kilima*) and tetraploid (*A. digitata*) species along a transect in the Baobab habitat of Tanzania and Kenya.

2.5 Molecular Markers

Molecular markers are useful tools for determining and understanding genetic variation and they provide an efficient means to link genotypic and phenotypic variation (Varshney et al., 2005). They are classified as: (i) PCR-based markers i.e. Amplified Fragment

Length Polymorphisms (AFLPs), Random Amplified Polymorphic DNAs (RAPDs), Inter Simple Sequence Repeats (ISSRs) and microsatellites also known as Simple Sequence Repeats (SSRs), (ii) hybridization based markers e.g. Restriction Fragment Length Polymorphisms (RFLPs), and (iii) sequence based markers i.e. Single Nucleotide Polymorphisms (SNPs) (Kalia, Rai, Kalia, Singh, & Dhawan, 2011) as shown in Table 1.

Table 1: Important features of different types of molecular markers (Kalia et al., 2011).

Features	Molecular Markers			
	EST-SSRs	SSRs	RFLPs	RAPDs/AFLPs/ISSRs
Need for Sequence Data	Essential	Essential	Not Required	Not Required
Level of Polymorphism	Low	High	Low	Low-moderate
Dominance	Co-dominant	Co-dominant	Co-dominant	Dominant
Interspecific Transferability	High	Low-moderate	Moderate-high	Low-moderate
Utility in Marker Assisted Selection	High	High	Moderate-high	Low-moderate
Cost and Labor in Generation	Low	High	High	Low-moderate

2.5.1 Microsatellite Markers

Hamada et al., (1982) was the first to demonstrate the existence of microsatellites in different eukaryotes ranging from yeasts to vertebrates. Following studies by Tautz & Renz., (1984) abundance of microsatellites in plants and other numerous eukaryotes was confirmed. Simple Sequence Repeats (SSRs) are found in both coding and noncoding regions and they are widely distributed throughout the nuclear genome (Kalia et al., 2011). SSRs are also found in chloroplast (Provan et al., 2001) and mitochondrial (Soranzo et al., 1999) genomes.

Microsatellites have been found to cause genetic variations in plants. These genetic variations are important in facilitating plant breeding and conservation (Kalia et al., 2011).

They can also be termed as Simple Sequence Repeats (SSRs), Short Tandem Repeats {STRs (Edwards et al., 1991)} or Simple Sequence Length Polymorphisms {SSLP} (Tautz, 1989). They are a unique type of DNA motifs repeated in tandem (ranging in length from 2-5 base pairs) and are repeated typically 5-50 times throughout the genome. They have been classified depending upon their location in the genome, type of repeat unit and size. Depending on the number of nucleotides per repeat unit, they are classified as mono-, di-, tri-, tetra-, penta- or hexanucleotides. Microsatellites can also be grouped based on their location in the genome, as chloroplastic (cpSSR), nuclear (nuSSR) or mitochondrial (mtSSR) SSRs (Kalia et al., 2011).

Microsatellites have further been classified as simple perfect repeats, simple imperfect repeats, compound perfect or compound imperfect markers. Simple perfect microsatellites are tandem arrays of a single repeat sequence (e.g., [AAG]n). Simple imperfect arrays consist of one or more repeat units made of different lengths (e.g., [AAC]n[ACT] [AAC]n+1). Compound perfect arrays are mainly composed of two or more different repeat motifs of equal length (e.g., [AGG]n[AATC]n) while compound-imperfect motifs are interrupted by one or more repeats of different length (e.g., [GGAT]n[ACT][GTAA]n+1) (Wang et al., 2009).

Table 2: Classification of Microsatellites (Kalia et al., 2011)**.**

(A) Based on the number of nucleotides per repeat	Mononucleotide (A)n
	Dinucleotide (CA)n
	Trinucleotide (CGT)n

	Tetranucleotide (CAGA)n
	Pentanucleotide (AAATT)n
	Hexanucleotide (CTTTAA)n (n = number of variables)
(B) Based on the arrangement of nucleotides in the repeat motifs	Pure or perfect or simple perfect (CA)n
	Simple imperfect (AAC)n ACT (AAC)n + 1
	Compound or simple compound (CA)n (GA)n
	Interrupted or imperfect or compound imperfect(CCA)n TT (CGA)n + 1
(C) Based on location of SSRs in the genome	Nuclear (nuSSRs)
	Chloroplastic (cpSSRs)
	Mitochondrial (mtSSRs)

Microsatellites have also been found to be ubiquitous in prokaryotes and eukaryotes (Gurarie et al., 2000). They are codominant markers relatively small in size and can be amplified easily using Polymerase Chain Reaction (PCR). This is done using two unique oligonucleotide primers that flank the microsatellite defining the microsatellite locus. They have been found to exhibit Mendelian inheritancec and can therefore be used for linkage analysis (Gupta et al., 1996). These features have made them to be successfully applied in a wide range of applied and fundamental fields of biology and medicine. This includes population and conservation genetics, forensics, genetic mapping, construction of linkage maps, marker assisted selections and backcrosses, molecular epidemiology, parasitology and genetic dissection of complex traits. Due to the variation in the number of repeats, high levels of allele polymorphism are demonstrated. Although they are considered to be selectively neutral markers, they contribute to chromatin organization, DNA structure, gene expression, regulation of DNA recombination, cell cycle dynamics, transcription and translation (Chistiakov et al., 2006).

SSR markers have gained considerable importance in plant breeding and genetics. This is due to their many desirable attributes. They include: co-dominant inheritance, relative abundance, hyper variability, multiallelic nature, reproducibility, extensive genome coverage, amenability to automation, chromosome specific location and high throughput genotyping (Kalia et al., 2011). Microsatellites can be developed from libraries enriched for specific microsatellites or directly from genomic DNA libraries. They can also be found by searching public databases for example European Molecular Biology Laboratory (EMBL) and GenBank. Cross-species transferability and random PCR amplification of genomic DNA (e.g. RAPDS) can also be used to generate microsatellites. Currently, Expressed Sequence Tag (EST) databases are a vital source of candidate genes. This is because they can generate markers directly associated with the desired trait and may be transferable in close relative genera (Kalia et al., 2011).

2.6 Phylogenetic inference of the African *Adansonia* species

The study and analysis of ribosomal DNA (rDNA) sequences and genes in living organisms can be used to provide a wealth of information about phylogenetic relationships and molecular evolution. These studies have been used to infer phylogenetic history among the basal lineages of life to relationships among closely related species and populations. The rDNA is widely utilized because of its systematic versatility.

It has several beneficial features which include; numerous rates of evolution both among and within rDNA genes, presence of many copies of the rDNA sequences per genome and the pattern of concerted evolution occurring among repeated copies (Hillis and Dixon, 1991). The rapid concerted evolution within and between loci promote its usefulness for phylogeny reconstruction (Baldwin, 1992). This evolution is through gene conversion and unequal crossing-over. It promotes intragenomic uniformity of repeat-units, sometimes even between nrDNA loci on non-homologous chromosomes (Wendel et al., 1995; Hillis et al., 1991), and ensures accurate reconstruction of species relationships from the

sequences. Sexual recombination and concerted evolution may also promote nrDNA uniformity within interbreeding populations and this would minimize the need for intrapopulational sampling while carrying out phylogenetic studies (Baldwin et al., 1995).

In the eukaryotic nuclear genome there are hundreds to thousands of copies of rDNA genes in tandem arrays. This high copy number promotes amplification, cloning, detection and sequencing of nuclear rDNA. Each of these arrays consists of regions that code for 3 subunit genes. They are the 18S, 5.8S and 28S which are separated by ITS1 and ITS2, the two internal transcribed spacers, and at the 5' end, an external transcribed spacer (ETS) located upstream of the 18S gene (Hillis and Dixon, 1991) as shown in Figure 2.

The internal transcribed spacers (ITS1 and ITS2) of nuclear rDNA have widely been utilized as sources of informative variation for inter-specific-/intergeneric-level analysis of phylogeny among angiosperms and other eukaryotes (Hershkovitz & Zimmer, 1996). The ITS sequences are also widely used for reconstruction of phylogenies in plants (Sang et al., 1995). In addition, variation in the spacer regions has been used to study hybridization, identify species or strains and as markers in population genetic studies (Hillis and Dixon., 1991).
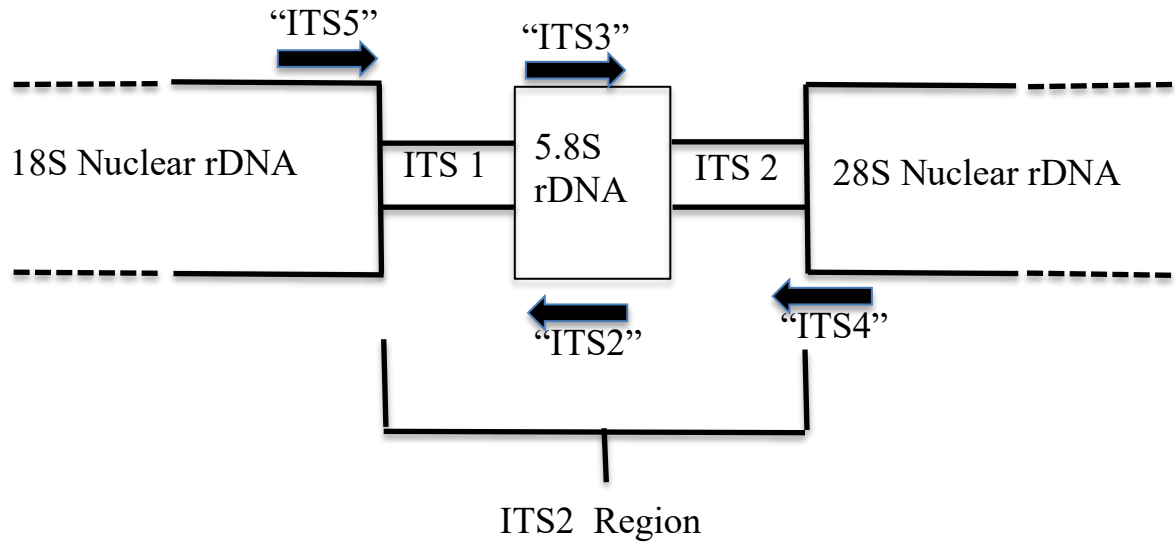
Figure 2: Organization of the Internal Transcribed Spacer region (Baldwin., 1992) showing ITS 1 and ITS 2 regions separated by the 5.8S ribosomal subunit, situated between the 18S and 28S nuclear rDNA. Arrows indicate approximate positions of primers designed by *(White et al., 1990)* used to amplify the ITS 1 and ITS 2 region.

The ITS region is conservative in length and small in plants with generally few restriction sites found within it (Baldwin, 1992). It is noncoding and is therefore a rapidly evolving region. It is easy to amplify using PCR as it is present in multiple copies throughout the ribosomal genes (Wagener et al., 2006). PCR amplification of the two internal transcribed spacers is facilitated by conserved flanking regions of the 18S, 5.8S and 28S genes (Hillis and Dixon., 1991). The small size of the ITS region (<700bp) and highly conserved sequences make it easy to amplify this region especially when using universal eukaryotic primers that were designed by White et al., (1990).

The ITS region is part of the transcriptional unit of nuclear ribosomal DNA. ITS-1 and ITS-2 function, in part, in the maturation of nuclear rRNAs. In yeast (*Saccharomyces cerevisiae), in vivo* mutational analysis indicates that deletions of certain regions within ITS-1 can inhibit the production of mature small and large subunit rRNAs. Certain point

mutations or deletions in ITS-2 prevent/reduce the processing of large subunit rRNAs (Baldwin et al., 1995). Variation between ITS sequences have been reported to be due to point mutations. Few of the sites are affected by deletions (indels) or insertions among the sequences that are similar enough retaining sufficient signal for phylogenetic analysis. Sequence alignment is therefore unambiguous, within these limits, except in small regions of lower structural constraint. Data sets combined from both spacers when analyzed phylogenetically, yield trees with great resolution and internal support than if analyses were based on either spacer alone. ITS characters have improved the understanding of angiosperm phylogeny by: resolving conflicts found between other data sets, providing direct evidence of reticulate evolution especially when concerted evolution doesn't act across repeat units contributed by different parental species, confirming earlier findings and improving resolution of species relationships (Baldwin et al., 1995).

CHAPTER THREE
MATERIALS AND METHODS

3.1 Study areas and sampling

The study was performed on Baobab tree populations established in the Eastern slopes of Mt. Kilimanjaro (Tsavo West National Park) and the South Kenyan Coast. Eight populations in total were sampled namely: From Kenya, Voi (9 samples), Kibwezi (15), Machinery (5), Mtito (10) and Malindi (6). From Tanzania, Moshi (5), Kondo (8) and Same (4). The coordinates, altitudes and sample number allocations are described in Appendix 1. These populations were selected based on their proximity to the Kenya/Tanzania boundary which was the region of interest and also as a replication of the work done by Pettigrew et al., (2012).

Thirty nine fruit samples previously collected from the Kenyan populations were available. Leaf sampling was therefore not required. From the Tanzanian populations, twenty three leaf samples were collected. A total of 62 samples from selected baobab trees were used. The sampled leaves were dried in an excess volume of silica-gel granules and stored at 4°C in sealed snap-top bags.

3.2 Seed germination

DNA was extracted directly from the Tanzanian leaf samples. The seeds from the Kenyan fruits were pretreated by mechanical scarification through nicking with a nail clipper. The seeds were then planted in polythene bags in the World Agroforestry Center (ICRAF) nursery and thereafter, DNA was extracted from the young leaves.

3.3 DNA isolation

For the samples that were planted in the nursery, young leaves were freshly picked and DNA extracted. For the Tanzanian samples, dried leaves were used. Genomic DNA was extracted from approximately 40 mg of the sample leaves using the Cetyl

trimethylammonium bromide (CTAB) method developed by (Doyle, 1987). Briefly, the leaves were ground using a Qiagen Tissue Lyser II (Hilden, Germany). A pinch of PVP (polyvinylpyrrolidone) was added during grinding as suggested by (Pettigrew et al., 2012). DNA was extracted using 800 µl of extraction buffer {2% CTAB, 100mM Tris-HCl (pH 8.0), 5M NaCl, 20mM EDTA (pH 8.0), 1% β-mercaptoethanol and 2% PVP}. The ground tissue and buffer mixture was incubated at 65°C for 45 minutes with occasional mixing to enhance cell wall degradation and release of the DNA.

An equal volume (800 µl) of Chloroform/Isoamylalcohol (CIAA solution; 24:1) was added to the mixture and the tubes were gently shaken to mix. This was done to eliminate cellular debris and proteins. The samples were centrifuged at 13200 rpm for 20 minutes using an Eppendorf 5415D (Hamburg, Germany) centrifuge. The supernatant was transferred to new labeled tubes and 0.7 volumes of ice cold Isopropanol added and incubated at 20 ˚C overnight to precipitate the nucleic acids. The samples were centrifuged at 13200 rpm for 20 minutes to pellet the DNA. The supernatant was carefully discarded not to dislodge the pellet. The pellets were washed twice using 70% ethanol centrifuging each time at 13200 rpm for 5 minutes. The tubes were then air-dried on a clean paper towel for 20-30 minutes. The DNA pellets were resuspended in 100 µl TE and 4 µl of RNase was added to degrade RNA. The samples were incubated for 30 minutes at 37 ˚C in a water bath before storage at 4 ˚C.

3.4 DNA quantification and quality analysis

DNA quantification and quality analysis was carried out by electrophoresis on a 0.8% agarose gel stained with 5 µl of GelRed (Biotium Inc. USA). Volumes of 3 µl of sample and 3 µl of loading dye (bromophenol blue, xylene cyanol and sucrose) were mixed gently by pipetting up and down before loading onto the agarose gel. The gel was run at 100V for 1 hour in 0.5X TBE buffer (0.1M Tris base, 0.1M boric acid and 0.02M EDTA; pH 8.0). One microliter of a 40 ng/µl standard molecular weight marker (Lambda DNA,

Invitrogen, USA) mixed with 2 µl of loading dye was also loaded. The gels were viewed under UV illumination at 312 nm and photographed using a gel documentation system (Vilber Lourmat TFX gel, France).

The concentration and purity of the DNA was measured using a Qubit® 2.0 Fluorometer (Invitrogen, USA) and NanoDrop 2000 spectrophotometer (Thermo Scientific, USA) respectively. Sample purity ratios (260/280 nm and 260/230 nm) were read. The ratio of absorbance at 260 nm and 280 nm was used to assess the purity of the DNA. A ratio of ~1.8 was accepted as "pure" for DNA. The DNA was then diluted to a concentration of 30 ng/µl and stored at 4 ˚C until use.

3.5 Primer selection and optimization

For the microsatellite analysis, 18 loci were selected and 18 SSR primers developed by Larsen et al., (2009) were used (Table 3 & 4). The markers were found to be codominant and therefore suitable for molecular ecological studies and appropriate for studying gene flow and distribution dynamics. The motif sizes were also considered as they were important in ensuring accurate binning during the allele scoring. The primers were named Primer 1-18 according to their Loci names Ad01-Ad18 as shown in Table 3.

Table 3: The selected SSR loci listed with GenBank accession numbers, forward and reverse primer sequences, motif sizes and expected sizes of amplicons (Larsen et al., 2009).

Locus Name	Primer sequences 5'-3'		Motif	Size Range (bp)
	F	R		
Ad01	CATTGCCAGGAATGCTTTTGC	GGATTGCCAGGTCTACTAC	$(AG)_{19}$	96-125
Ad02	TGCTGACTAGCAGTTTCCTATG	TCAGATGCCAAACATTCACACC	$(TC)_{15}$	267-295
Ad03	GGATCAAATTATGGTTAAGGC	CCAATTTTGAGCCAATTCTCA	$(GA)_{21}$	143-175
Ad04	GTTGCTTGTGTGCTTACCC	CATCCCTCTCCCCATTCC	$(CT)_{20}$	173-236
Ad05	CTCAACAAGGTTCGGATGTCGTATG	GTCTGCCGGGTGTTTTGCATG	$(CA)_{12}(CT)_{12}$	295-319
Ad06	TGCATCAGCTTTCACTCCAGAC	GCCACCCATAAAACCCAATCC	$(TC)_{19}$	129-154
Ad07	TAGAAAATTAGCAGATAAGTGC	GATTTCGGTGATATGTTGTAG	$(AG)_{18}$	159-193
Ad08	TCTAAAGCCTGTAAGGAAAAATGGG	TTCTCCGTTCACTCTGTACTTCC	$(GAA)_{14}$	267-296
Ad09	TACCACTTCTCCAGATGCTAC	ACTGGCTAGAGATGCGTTG	$(AAG)_{11}$	190-209
Ad10	GCAGCTTGCCTCGTCATATA	CCAATGGCAATGTGTCTGACG	$(CT)_6CC(CT)_$	216-262
Ad11	ATCAGCCATTCTGCATACCTGC	TAGGCACAAAACTGAGATGCACAG	$(CA)_{13}(AT)_6$	118-181
Ad12	GCTTGTCAAGCAATTCCCC	ACTTTGTCCCACCTGTTTCTC	$(AG)_{16}$	162-175
Ad13	CCCCACTTCAGATCAAGTAAGTC	GCTGTATTTCTGAGCCTGAGAAG	$(AC)_{14}$	305-330
Ad14	CTTGATTGGAATACGGGAAATGGAG	CCAAACCAATTGGACTTTGACCTTC	$(AC)_{13}$	170-191
Ad15	TGAAGAGACAAAGCAAGAAG	CATGACATCTCCTTGAACC	$(GAA)_{14}$	130-161
Ad16	TGCATCAGCTTTCACTCCAGAC	GCCACCCATAAAACCCAATCC	$(TA)_5(TG)_{19}$	219-254
Ad17	GCGCCTTAGAAAGGACTTGTTAGAG	GCCAACAGCCTTAGTAGTCCAAG	$(AC)_{14}$	174-215
Ad18	ACCGCTTCCGTTCTCATTCC	ACCACCACTACACCGTCATTG	$(TG)_{17}$	257-291

The forward strands of the SSR primers were fluorescently labelled with the following dyes: VIC (Green), NED (Black), PET (Red) and FAM (Blue) from Applied Biosystems, USA. Several Gradient PCR's were run to optimize the primers based on the annealing temperatures published. Sixteen of the 18 primers were selected for the final analysis as shown in Table 4.

Table 4: Sixteen SSR primers used for Microsatellite analysis.

Locus	Forward Primer	Reverse Primer	Tm (°C)
Ad01	CATTGCCAGGAATGCTTTTGC	GGATTGCCAGGTCTACTAC	52
Ad02	TGCTGACTAGCAGTTTCCTATG	TCAGATGCCAAACATTCACACC	53
Ad04	GTTGCTTGTGTGCTTACCC	CATCCCTCTCCCCATTCC	58
Ad05	CTCAACAAGGTTCGGATGTCGTATG	GTCTGCCGGGTGTTTTGCATG	58
Ad06	TGCATCAGCTTTCACTCCAGAC	GCCACCCATAAAACCCAATCC	55
Ad07	TAGAAAATTAGCAGATAAGTGC	GATTTCGGTGATATGTTGTAG	47
Ad08	TCTAAAGCCTGTAAGGAAAAATGGG	TTCTCCGTTCACTCTGTACTTCC	54
Ad09	TACCACTTCTCCAGATGCTAC	ACTGGCTAGAGATGCGTTG	52
Ad10	GCAGCTTGCCTCGTCATATA	CCAATGGCAATGTGTCTGACG	52
Ad11	ATCAGCCATTCTGCATACCTGC	TAGGCACAAAACTGAGATGCACAG	55
Ad12	GCTTGTCAAGCAATTCCCC	ACTTTGTCCCACCTGTTTCTC	51
Ad14	CTTGATTGGAATACGGGAAATGGAG	CCAAACCAATTGGACTTTGACCTTC	56
Ad15	TGAAGAGACAAAGCAAGAAG	CATGACATCTCCTTGAACC	48
Ad16	TGCATCAGCTTTCACTCCAGAC	GCCACCCATAAAACCCAATCC	55
Ad17	GCGCCTTAGAAAGGACTTGTTAGAG	GCCAACAGCCTTAGTAGTCCAAG	58
Ad18	ACCGCTTCCGTTCTCATTCC	ACCACCACTACACCGTCATTG	54

3.6 ITS Primers used to amplify the ITS region

The nuclear ribosomal DNA region including the internal transcribed spacers ITS-1 and ITS-2 were amplified using 4 primers (ITS LEU, ITS 4, ITS S3R and ITS S2F) described by Baum et al., 1998; Taberlet et al., 1991; Kress et al., 2005 respectively as shown in Table 5 and produced by Applied Biosystems, USA.

Table 5: Primers used to amplify the ITS region

Region	Primers	Annealing Temperature (˚C)	Sequences
ITS	ITSLEU Reverse	56	5' GTCCACTGAACCTTATCATTTAG 3'
	ITS4 Forward		5' TCCTTCCGCTTATTGATATGC 3'
	ITS S3R Reverse	56	5' GACGCTTCTCCAGACTACAAT 3'
	ITSS2F Forward		5' ATGCGATACTTGGTGTGAAT 3'

3.7 PCR amplification of SSRs

For the SSR amplification, a 20 µl PCR master mix was prepared with the following specifications per reaction: 4 µl of 5X MyTaq™ Reaction Buffer, (Bioline, U.K) with pre-optimized concentrations of dNTPs, $MgCl_2$ stabilizers and enhancers, 2 µl of 2 µM fluorescent labelled forward primer, 2 µl of 2 µM reverse primer, 0.4 µl of MyTaq™ DNA Polymerase (Bioline, U.K) and 7.6 µl of distilled water. For every reaction, 4 µl (20 ng/µl) of sample DNA was added.

PCR was carried out on two thermocyclers. A Veriti 96-well Thermal Cycler (Applied Biosystems, USA) for the 96 well plates and a Gene Amp PCR System 9700 (Applied Biosystems, USA) for the 384 well plates. Touchdown PCR was carried out for the primers with similar annealing temperatures on a 384 well plate and Conventional PCR

was carried out for the primers with different annealing temperatures on a 96 well plate as shown on Table 4. Different PCR protocols were used as Touchdown PCR requires a higher initial annealing temperature than the optimal Tm which is gradually reduced over subsequent cycles until the touchdown or Tm temperature is reached.

The following amplification protocol was utilized for the Touchdown PCR: An initial denaturation at 95 ˚C for 3 minutes followed by 10 cycles of denaturation at 95 ˚C for 15 seconds, annealing at a slightly higher temperature than the optimal annealing temperature for 30 seconds and an extension at 72 ˚C for 15 seconds. The next 25 cycles followed a denaturation at 95 ˚C for 15 seconds, annealing at gradually reduced temperatures over subsequent cycles for 30 seconds (reduction by 3 ˚C) and an extension at 72 ˚C for 15 seconds. The holding temperature was 15 ˚C for ∞ time.

The PCR amplification protocol for the PCR carried out on a 96 well plate consisted of an initial denaturation at 95 ˚C for 3 minutes followed by 35 cycles of the following profile: Denaturation at 95˚ for 15 seconds, annealing at the optimized annealing temperature (Table 4) for 30 seconds and an extension at 72 ˚C for 1 minute. A final extension followed at 72 ˚C for 10 minutes and a holding temperature at 15 ˚C for ∞ time

3.8 PCR amplification of ITS region

For the ITS amplification, PCR was carried out using the following specifications per reaction: 4 µl of 5X MyTaq™ Reaction Buffer, (Bioline, U.K) with pre-optimized concentrations of dNTPs, $MgCl_2$ stabilizers and enhancers, 2 µl of 2 µM each of forward and reverse primers, 0.4 µl of MyTaq™ DNA Polymerase (Bioline, U.K) and 7.6 µl of distilled water. For every reaction, 4 µl of sample was added. The total volume of the reaction was 20 µl.

The PCR amplification protocol consisted of an initial denaturation at 95 ˚C for 3 minutes followed by 35 cycles of the following profile: Denaturation at 95 ˚C for 15 seconds, annealing at 56 ˚C for 30 seconds and an extension at 72 ˚C for 1 minute. A final extension followed at 72 ˚C for 10 minutes and a holding temperature at 15 ˚C for ∞ time.

3.9 Quantification of PCR products

Quantification of amplification products for both SSR and ITS amplifications was carried out as described in section 3.3 except for a higher percent gel (2 %) and smaller volumes of product and dye (2 µl each) were used.

3.10 Purification of PCR products

The ITS amplification products were purified for sequencing. Purification was done to effectively remove dNTPs, unincorporated nucleotides, primers, enzymes and salts from the PCR reaction. This was done using a GeneJet PCR Purification kit (Thermo Scientific, USA) following the manufacturer's instructions. The samples were then diluted to a concentration of 50 ng/µl and submitted to Segolip Labs ILRI-Kenya for sequence analysis.

3.11 Preparation of SSR amplification products for Capillary Electrophoresis

Into a 2 ml Eppendorf tube, 800 µl of HiDi™ Formamide (Applied Biosystems, UK) and 12 µl of 500 LIZ TM Size Standard (GeneScan TM, Applied Biosystems, UK) were added and the mixture vortexed thoroughly. From the Hi-Di/Liz mixture, 8 µl was aliquoted into each well of the 96-well plate. HiDi is a highly deionized formamide used to resuspend samples, in order to prevent reannealing of complementary strands before they are electrokinetically injected on capillary electrophoresis systems. The LIZ Size Standard is a dye-labelled size standard used for reproducible sizing of fragment analysis data. It sizes DNA fragments in the 35-500 nucleotides range and provides 16 single stranded labeled fragments of: 35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, 350, 400, 490 and 500

nucleotides. Different PCR products from different SSR markers were co-loaded. Each 96-well plate was co-loaded with 3 different SSR marker products each labelled with a different fluorescent dye. One microliter of each sample was added to the wells containing 8 µl of the HiDi/Liz mixture. The plates were denatured at 95 ˚C for 5 minutes and snap cooled in ice for 15 minutes. The plates were then sent to SegoLip Labs, ILRI –Kenya for capillary electrophoresis. The PCR products were size-fractioned using an ABI 3730 Capillary DNA sequencer (PE-Applied Biosystems, USA).

3.12 Microsatellite allele scoring and analysis

Scoring and determination of the size of the alleles was done using GeneMapper® version 4.0 (Thermo Fischer Scientific, USA) and the GS500LIZ TM internal size standard. The data was analyzed as a binary data matrix and the SSR marker treated as dominant markers. The presence or absence of each PCR amplification product was scored as 1 or 0 respectively and the data matrix was generated using ALS Binary (Prasanth and Chandra., 1997). Allelobin (Prasanth et al., 1997) was used to classify the observed microsatellite allele sizes into representative discrete alleles. It was also used to determine the accuracy of allele sizing and the quality of the automated binning. This was represented by the Quality Index (QI) of each marker.

3.13 Analysis of Molecular Variance (AMOVA)

In order to assess the distribution of variation within and among the populations, an analysis of molecular variance was carried out. GenAlex version 6.41 (Peakall and Smouse, 2006) was utilized. An output pairwise PhiPT matrix was generated at 999 permutations. The PhiPT and P values were also determined. The PhiPT values were used to attach significance to the variation estimates and summarize the degree of differentiation among the populations. An analysis of private alleles within the populations was also carried out.

3.14 Principal Coordinate Analysis (PCoA)

A Principal Coordinate Analysis was carried out to visualize the genetic relationship and clustering among all individuals. GenAlex version 6.41 (Peakall and Smouse., 2006) was used. The principal coordinates were generated based on genetic distance through a covariance matrix with data standardization. The populations were color coded for easier visualization.

3.15 Cluster analysis

The cluster analysis was carried out using DARwin6 software. The raw data was used to generate a dissimilarity matrix using the Jaccard similarity coefficient at 100 bootstraps. During the analysis, '0' was taken as 'Absence' and '1' 'Presence' of an allele at all loci in each individual. The dissimilarity matrix was then utilized in generating the Unweighted Neighbor Joining tree and an Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendogram. It was also used to carry out the factorial analysis.

3.16 Measures of genetic diversity

To determine the genetic diversity of the populations, the average heterozygosity and percentage polymorphic loci at 99 % were calculated using Tools for Population Genetic Analyses (TFPGA) version 1.3 (Miller, 1997). The allele frequencies were estimated based on the frequency of the null genotype and the Hardy-Weinberg equilibrium was assumed. An unweighted pair-group method with arithmetic average (UPGMA) dendogram was also constructed. A total of 1000 permutations were computed and confidence levels of the dendogram estimated.

3.17 Polymorphic Information Content (PIC) value determination

Powermarker version 3.25 (Liu & Muse, 2005) was utilized to analyze the data generated by GeneMapper. Summary statistics of major allele frequency, gene diversity and PIC values were generated. The level of polymorphism among the genotypes was determined for each of 16 SSR primers.

3.18 Phylogenetic analysis of the African *Adansonia* species

The raw sequences were edited and the forward and reverse strands aligned to build contiguous sequences using BioEdit version 7.0 (Hall, 1999). Outgroup sequences were obtained from GenBank. They included sequences of the species *Cavanillesia platanifolia* and *Bombax buonopozense.* Sequences of at least one representative of 6 species in the *Adansonia* genus were also included in the analysis. They included *A. rubrostipa, A. grandidieri, A. suarezensis, A. madagascariensis, A. perrieri and A. gibbosa.* Their accessions were retrieved from GenBank. Multiple sequence alignments of all sequences were created using MUSCLE plugin in MEGA6 (Tamura et al., 2013). Regions where the alignment was ambiguous were removed and the gaps were treated as unknown. The molecular phylogeny and generation of phylogenetic trees was done using Maximum Likelihood in MEGA6. Parsimony and bootstrap analyses were carried out with 1000 bootstrap iterations each with 100 heuristic search replicates and random addition of Tree Bisection and Reconnection (TBR) branch swapping and taxa.

CHAPTER FOUR

RESULTS

4.0 DNA Extraction

Using the CTAB protocol all the samples that were extracted produced clean, good quality and high molecular weight DNA. The ratio of absorbance at 260 nm and 280 nm was 1.8 indicating pure DNA. The samples were run on 0.8% agarose gel stained with 5 µl of GelRed (Biotium Inc.) against a 40 ng/µl standard molecular weight marker (Lambda DNA, Invitrogen) labelled L in Figure 3. All the samples were labelled 1- 62 as indicated in Appendix 1.
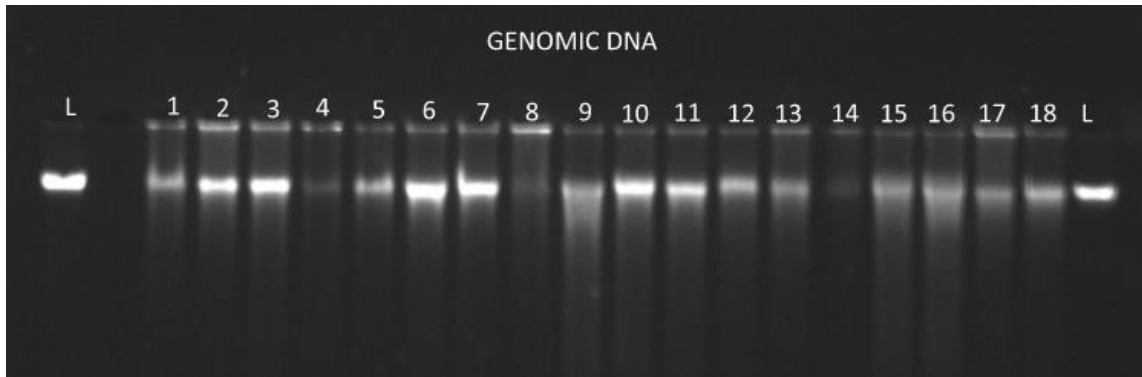
Figure 3: An agarose gel (0.8% w/v) image of extracted genomic DNA from sampled *Adansonia spp* leaves. L represents the Lambda DNA (Invitrogen). Some of the samples were extracted in multiple tubes shown in Appendix 2.

4.1 SSR Amplification

Eighteen SSR markers were optimized using touch-down PCR. Primers Ad03 and Ad13 did not give any amplicons. The remaining 16 primers amplified successfully with the expected product sizes and the results were used for microsatellite analysis. Figure 4 represents a gel image of one of the successfully amplified SSR primers, Primer 2 with

the expected band size of 267-295bp. Agarose gel images of all primers showing the amplification products are listed in Appendix 3-16.
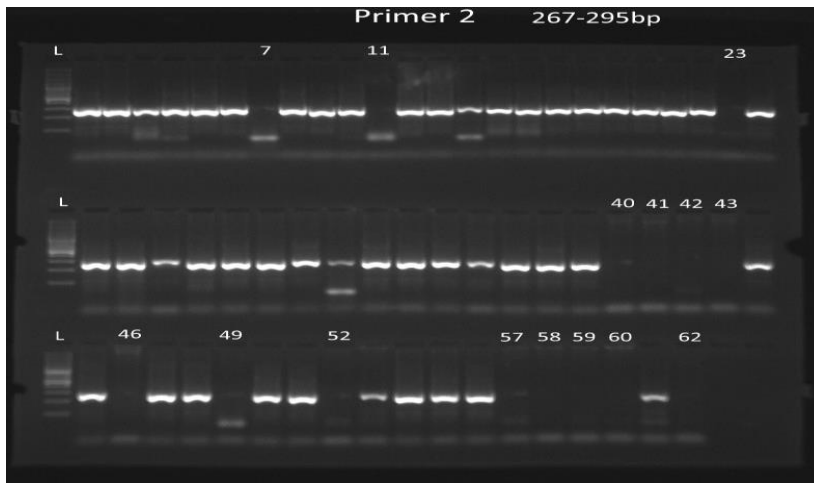
Figure 4: Agarose gel (2% w/v) image showing the amplification products of marker Ad02 with an expected product size of 267 – 295bp. L represents the Lambda DNA (Invitrogen). The numbered samples did not amplify.

4.2 Allele Scoring

The amplified fragments were successfully separated through capillary electrophoresis and allele sizes scored. The produced peaks showed the allele sizes at each loci. The analyzed samples had either two, three or four peaks in the loci amplified as shown in Figure 5.
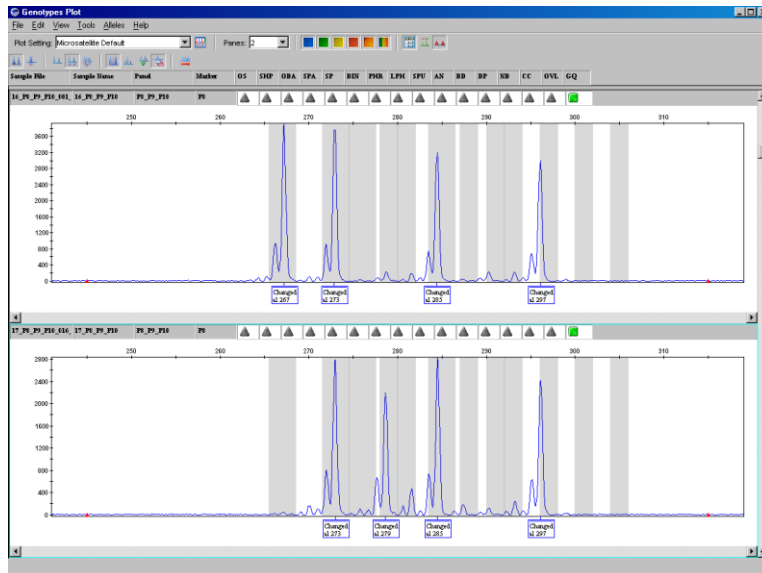
Figure 5: Screenshot of allele scored peaks showing allele sizes after amplification with Primer Ad08 that was fluorescently labelled with the FAM (Blue) dye.

4.3 Analysis of Molecular Variance

The AMOVA analysis showed low levels of genetic variation among populations with a P value of 0.001 and PhiPT value of 0.042 as shown in Table 6. The P value was used to determine the certainty of the associations. It provides the probability that a given result from a test is due to chance. PhiPT is a measure that allows intra-individual variation to be suppressed making it ideal to determine genetic differentiation between populations (Teixeira et al, 2014). It estimated that the genetic variation was partitioned with 4% among populations and 96% within populations as shown in Table 6.

Table 6: Summary of analysis of molecular variance (AMOVA) within and among *Adansoni*a populations (SS-Sum of squares, MS-Mean of squares, df- degree of freedom, PhiPT-Total Genetic Differentiation, Pvalue- 95% confidence interval).

Source	df	SS	MS	Est. Var.	%	PhiPT	Pvalue
Among Pops	7	312.570	44.653	1.471	4%	0.042	0.001
Within Pops	54	1813.301	33.580	33.580	96%		
Total	61	2125.871	78.233	35.050	100%		

4.4 Principal Coordinate Analysis

The PCoA was used to visualize the genetic relationship among all the individuals. The principal coordinates were calculated based on genetic distance (total number of allelic differences) between the samples using a covariance matrix with data standardization. The Kenyan populations as shown by Kibwezi, Voi, Mtito Andei and Machinery clustered close together while in the Tanzanian populations, Kondo clustered separately while the other Tanzanian populations overlapped with the remaining Kenyan populations as shown in Figure 6. Some general correlation was observed between clustering and geographic origin, indicating some degree of genetic structuring in the populations, although this factor is weakly supported statistically by the P value of 0.001.
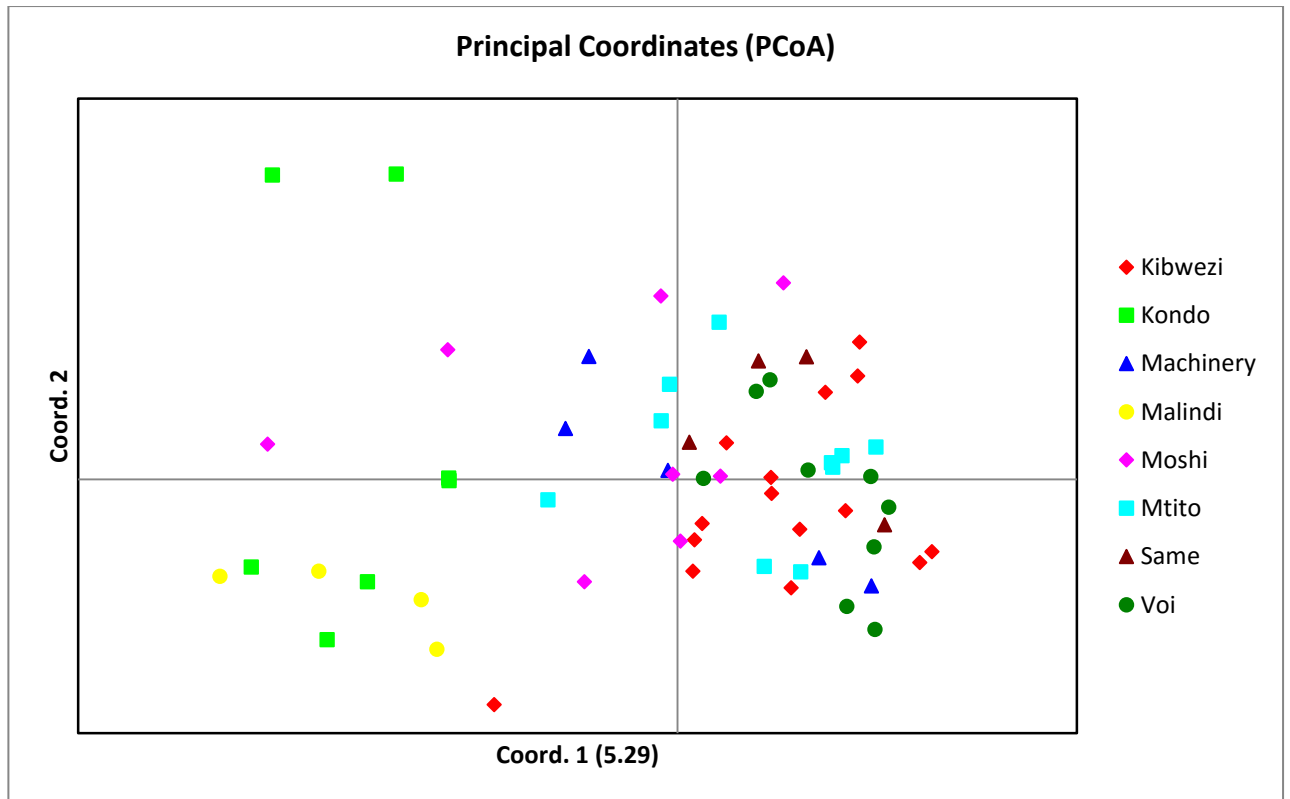
Figure 6: Principal coordinate analysis of 8 *Adansonia* populations based on genetic distances between all 62 samples. The Kenyan and Tanzanian populations clustered separately.

4.5 Cluster Analysis

The cluster analysis was carried out using DARwin6 software. The dissimilarity matrix was utilized in generating the UPGMA dendogram shown in Figure 7 and the neighbor joining tree shown in Figure 8. The UPGMA dendogram showed the grouping of the genotypes into 3 distinct clusters. All 3 clusters further grouped into sub groups. The neighbor joining radial tree also confirmed the 3 clusters. According to the dendogram, Cluster 1 (C1) mainly consisted of Kenyan populations i.e. Mtito Andei, Kibwezi, Voi and Machinery, Cluster 2 (C2) contained both Kenyan and Tanzanian populations while Cluster 3 (C3) consisted mainly of Malindi and Kondo populations from Kenya and Tanzania respectively.
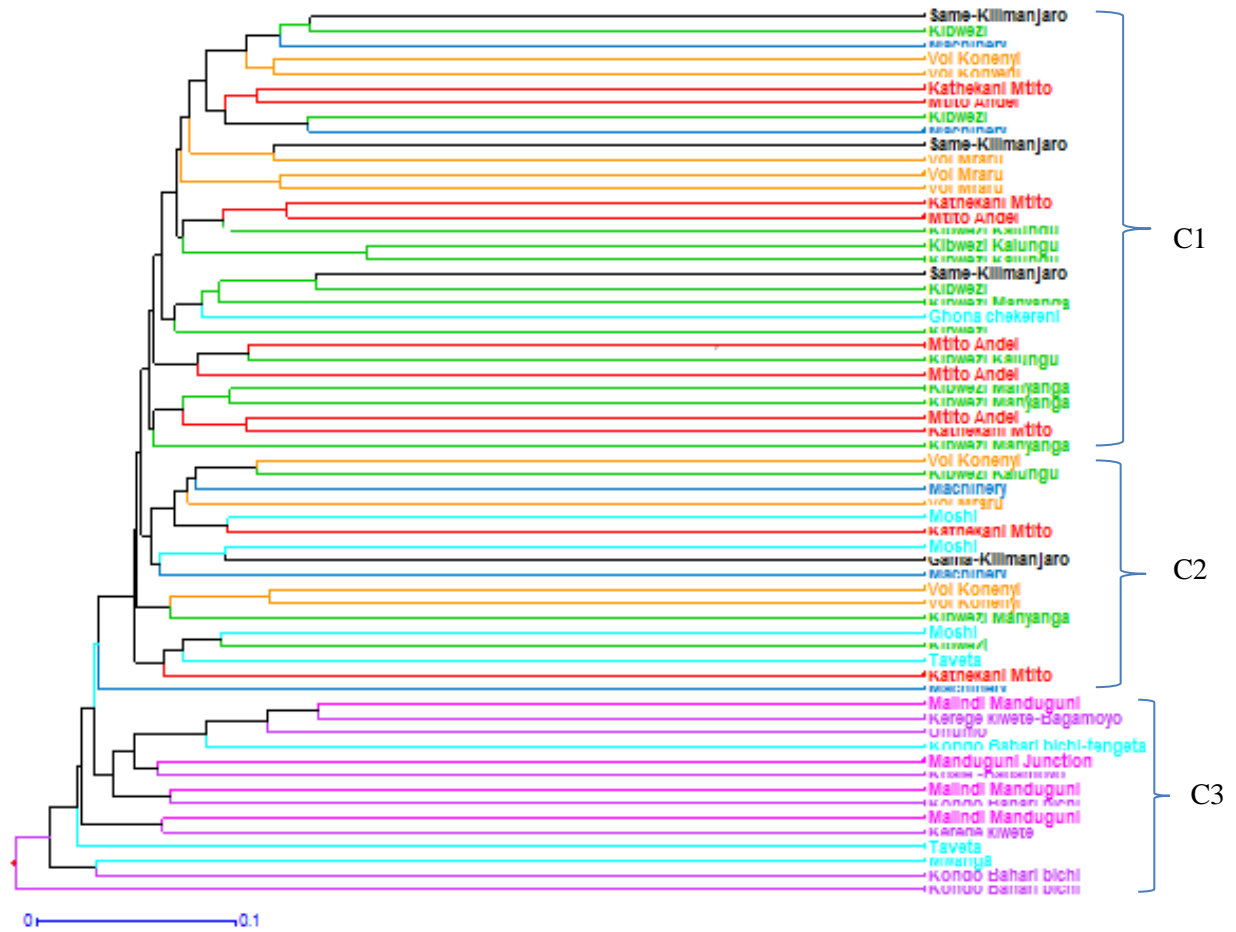
Figure 7: UPGMA dendogram computed for 62 individuals sampled across 8 baobab populations.

4.6 Neighbour Joining Radial Analysis

Using the unweighted neighbor joining method a radial tree was generated shown in Figure 8. This too showed 3 distinct clusters supporting the UPGMA dendogram generated by DARwin6.
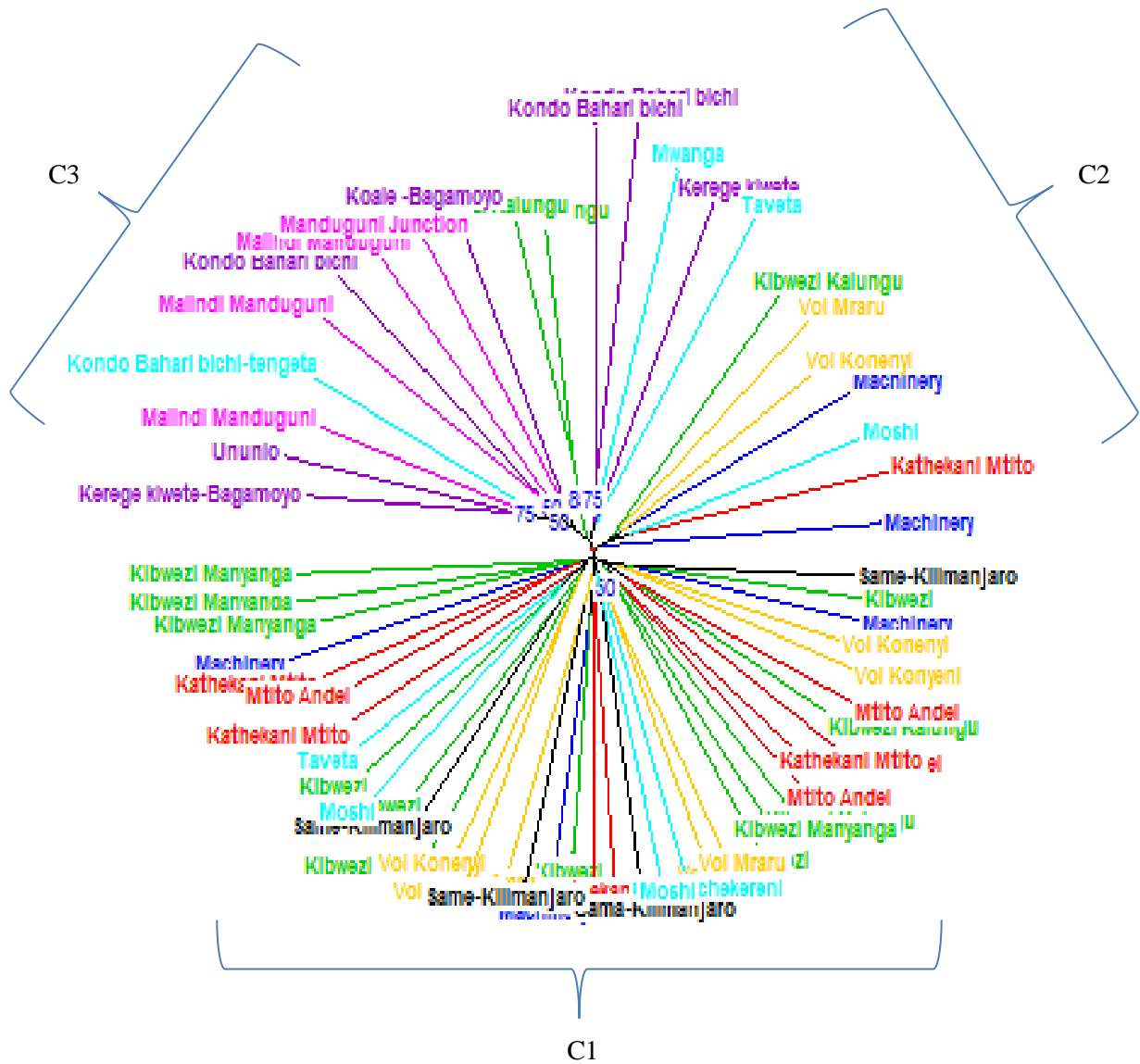
Figure 8: Neighbour joining radial tree showing 3 clusters (C1, C2 and C3). The tree was computed for 62 Baobab individuals sampled from 8 populations studied.

4.7 UPGMA dendogram generated by TFPGA

The UPGMA tree was constructed based on the genetic distance by TFPGA shown in Figure 9. It divided the populations into 2 clusters. These clusters divided the populations mainly based on their geographical positions and structure. Kondo and Malindi

populations being at the same elevation clustered together while the other group consisted of Kibwezi, Mtito Andei, Voi, Moshi, Machinery and Same.
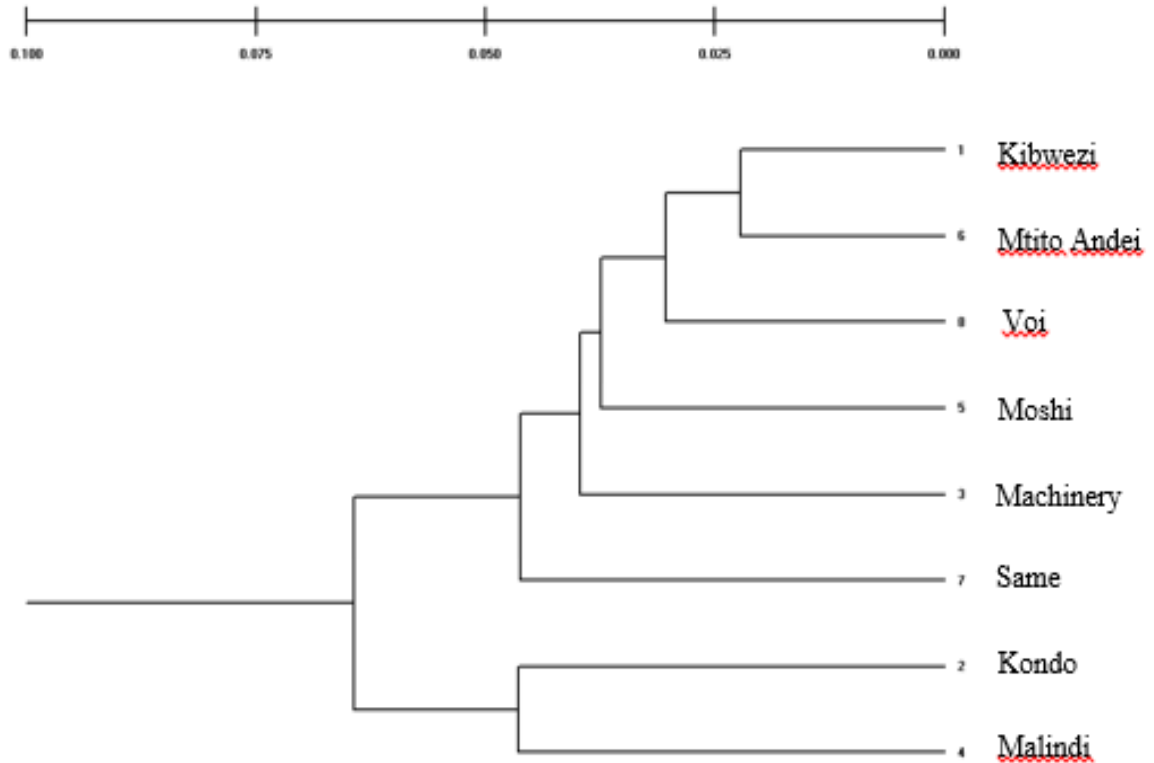
Figure 9: A UPGMA dendogram generated by TFPGA showing relationships among 8 populations of *Adansonia* species from the Kenya/Tanzania transect based on 335 loci amplified using 16 microsatellite markers.

4.8 Analysis of Band Patterns

Private alleles were found in all populations from both Kenya and Tanzania. Kondo population had the highest number of private alleles with 17 followed by Machinery, Moshi and Mtito populations with 9 private alleles as shown in Table 7. Kibwezi, Malindi and Same populations had 8, 6 and 2 private alleles respectively. Voi, with 1, had the

lowest number of private alleles. Kondo had the highest mean diversity of 0.194 with Same and Malindi having the lowest mean diversity of 0.141 and 0.140 respectively as shown in Figure 10.
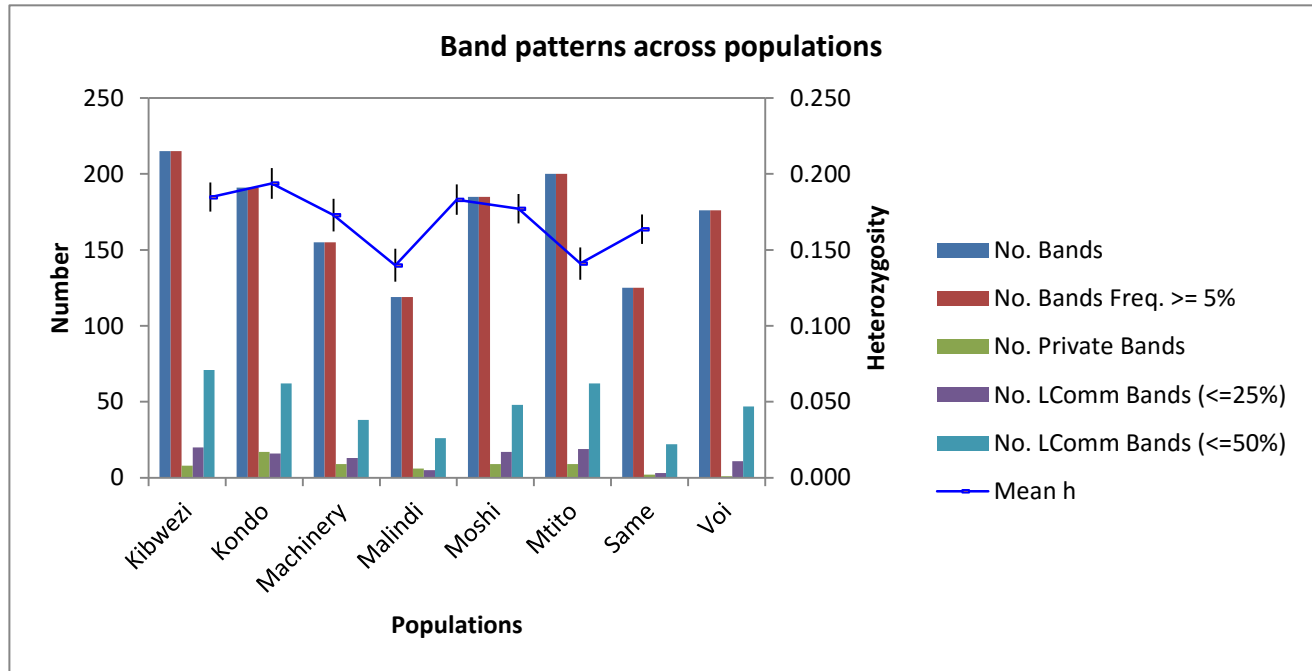
Figure 10: A bar graph showing number of private bands and the mean diversity of the populations.

Band Patterns for Binary Data by Populations								
Population	Kibwezi	Kondo	Machinery	Malindi	Moshi	Mtito	Same	Voi
No. Private Bands	8	17	9	6	9	9	2	1
Mean h	0.185	0.194	0.173	0.140	0.183	0.177	0.141	0.164

Table 7: Number of Private Bands and Mean Diversity of all the 8 populations

4.9 Genetic differentiation among populations

The genetic differentiation/variation between populations was determined using Pairwise PhiPT values showed in Table 8. This measure is ideal as it allows for the intra-individual variation to be suppressed. It is also efficient in comparing codominant and binary data with 999 permutations. The genetic differentiation coefficient of the 8 populations was 0.042 which showed that the genetic differentiation among populations was low. The Pairwise PhiPT genetic distances among populations ranged from 0.003 to 0.135. The genetic distance between Malindi and Voi populations was the highest at 0.135 while Machinery and Same populations had the lowest genetic distance at 0.003.

Table 8: Pairwise PhiPT genetic distances showing the genetic differentiation among the populations.

Binary Pairwise Population PhiPT Values								
Kibwezi	**Kondo**	**Machinery**	**Malindi**	**Moshi**	**Mtito**	**Same**	**Voi**	
0.000								**Kibwezi**
0.079	0.000							**Kondo**
0.013	0.047	0.000						**Machinery**
0.088	0.000	0.071	0.000					**Malindi**
0.045	0.022	0.000	0.065	0.000				**Moshi**
0.008	0.075	0.006	0.094	0.022	0.000			**Mtito**
0.023	0.069	0.003	0.122	0.000	0.000	0.000		**Same**
0.025	0.105	0.021	0.135	0.054	0.038	0.029	0.000	**Voi**

4.10 Measures of genetic diversity

Eight baobab populations were analyzed. The average heterozygosity (H) (Nei's genetic diversity) was based on 335 loci across the 8 populations. The H (average heterozygosity) values ranged from 0.1529 to 0.2736 as shown in Table 9. The percentage of polymorphic loci was based on a 99% criterion and it ranged from 34.0299 to 63.8806. It was observed that accessions with the highest average heterozygosity also had the highest percentage of polymorphic loci and the opposite was also true. Kibwezi had the highest value for average heterozygosity and highest % polymorphic loci while Malindi had the lowest values for both parameters. The Hardy-Weinberg Equilibrium was assumed and allele frequencies were estimated based on the square root of the frequency of the null genotype.

Table 9: Average heterozygosity (H) and % polymorphic loci of the population

Populations	Number of individuals	Average heterozygosity(H)	(99% criterion) Polymorphic loci
Kibwezi	15	0.2736	63.8806
Kondo	7	0.2653	57.0149
Machinery	5	0.2062	45.6716
Malindi	4	0.1529	34.0299
Moshi	8	0.2491	55.2239
Mtito	10	0.2576	58.8060
Same	4	0.1549	34.9254
Voi	9	0.2246	51.6418

4.11 Analysis using Power Marker

A total of 822 alleles were scored among the 8 populations using 16 markers. All the 16 SSR markers were found to be highly polymorphic among the populations with all PIC values being higher than 0.90. The loci Ad2, Ad6, Ad11 and Ad16 scored the highest allele number at 61, highest gene diversity at 0.9834 and highest PIC value at 0.9831, shown in Table 10. These loci showed the highest level of polymorphism. Ad9 and Ad10 loci scored the lowest allele number at 31. The mean PIC value was 0.9722.

Table 10: Measures of gene diversity, Allele frequency and Polymorphism information content using 16 markers with Powermarker.

Marker	Major.Allele.Frequency	Allele No	GeneDiversity	PIC
Ad1	0.0484	52.0000	0.9776	0.9771
Ad2	0.0323	61.0000	0.9834	0.9831
Ad4	0.0323	59.0000	0.9823	0.9820
Ad5	0.1129	49.0000	0.9693	0.9685
Ad6	0.0323	61.0000	0.9834	0.9831
Ad7	0.1129	45.0000	0.9610	0.9597
Ad8	0.0645	51.0000	0.9745	0.9739
Ad9	0.1129	31.0000	0.9448	0.9422
Ad10	0.0968	31.0000	0.9490	0.9467
Ad11	0.0323	61.0000	0.9834	0.9831
Ad12	0.0484	53.0000	0.9781	0.9777
Ad14	0.0806	47.0000	0.9693	0.9684
Ad15	0.0806	52.0000	0.9750	0.9745

Ad16	0.0323	61.0000	0.9834	0.9831
Ad17	0.0968	56.0000	0.9755	0.9750
Ad18	0.0484	52.0000	0.9776	0.9771
MEAN	0.0665	51.3750	0.9730	0.9722

4.12 Amplification of the ITS region

ITS specific primers (ITS S2F-S3R and ITS LEU-4) were used for amplification of the ITS region. The primer combinations that yielded results were ITS LEU/ITS4 and ITSS3R/ITSS2F. Upon optimization, amplification was successful at an annealing temperature of 56 ˚C. The amplification products are shown in Figure 11 and 12 where the ITS S2F-S3R generated bands ranging from 400-500bp and ITS LEU-4 bands ranged from 700-800bp.
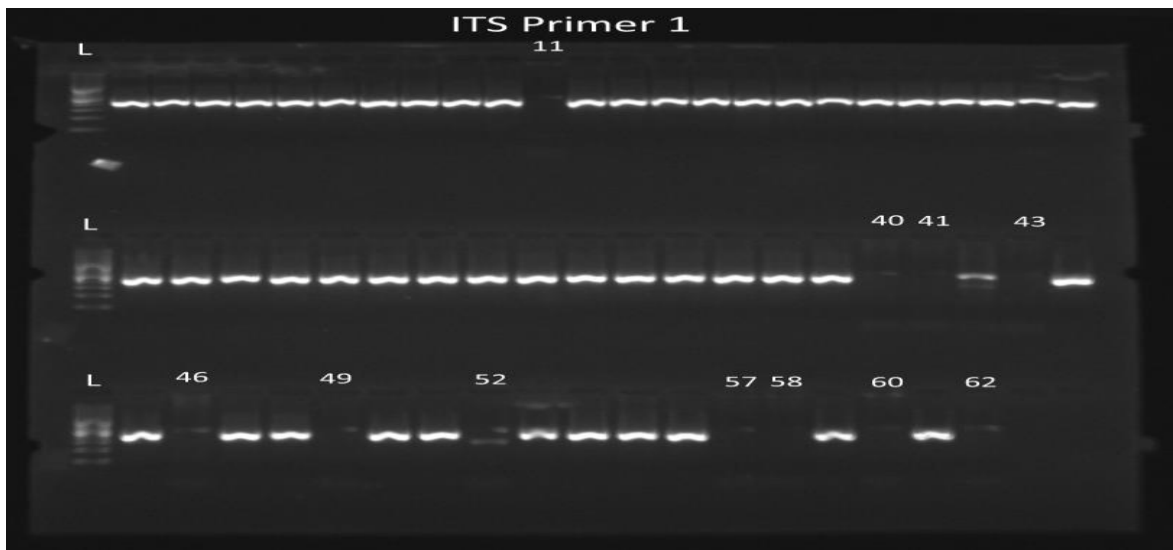
Figure 11: Agarose gel (2% w/v) image showing the amplification products of ITS S2F-S3R with generated bands ranging from 400bp-500bp. The numbered samples did not amplify. L represents the Lambda DNA (Invitrogen).
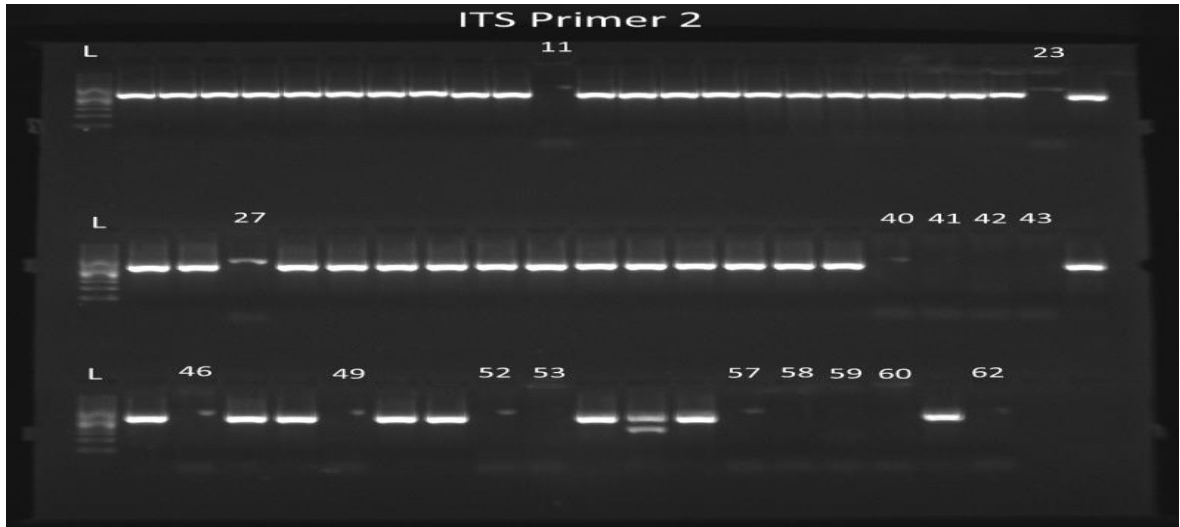
Figure 12: Agarose gel (2% w/v) image showing the amplification products of ITS LEU-4 with generated bands ranging from 700bp-800bp. The numbered samples did not amplify. L represents the Lambda DNA (Invitrogen).

4.13 Editing and alignment of consensus sequences

Sixty two samples were amplified but after sequence editing, 53 consensus sequences were generated for ITS LEU-4 and 52 consensus sequences for ITS S2F-S3R.

4.14 Alignment of Consensus Sequences

The ITS region that was analyzed had an aligned sequence length of 785bp. The individual samples sequences ranged from 768-779bp. Phylogenetic trees generated did not show any significant difference between *Adansonia digitata* and *Adansonia kilima* as shown in Figure 13 and 14.

4.15 Phylogenetic trees generated from ITS sequenced amplicons

Figure 13: Phylogenetic tree generated from amplification with ITS primer LEU/4

Figure 14: Phylogenetic tree generated from amplification with ITS primer S2F/S3R

CHAPTER FIVE

DISCUSSION

5.1 Genetic diversity and level of polymorphism

Genetic variation is a key requirement for adaptation and evolution and is, therefore, significant for the survival of plant populations (May, 1994; Wiehle et al., 2014; Mace et al., 2003; Jump et al.,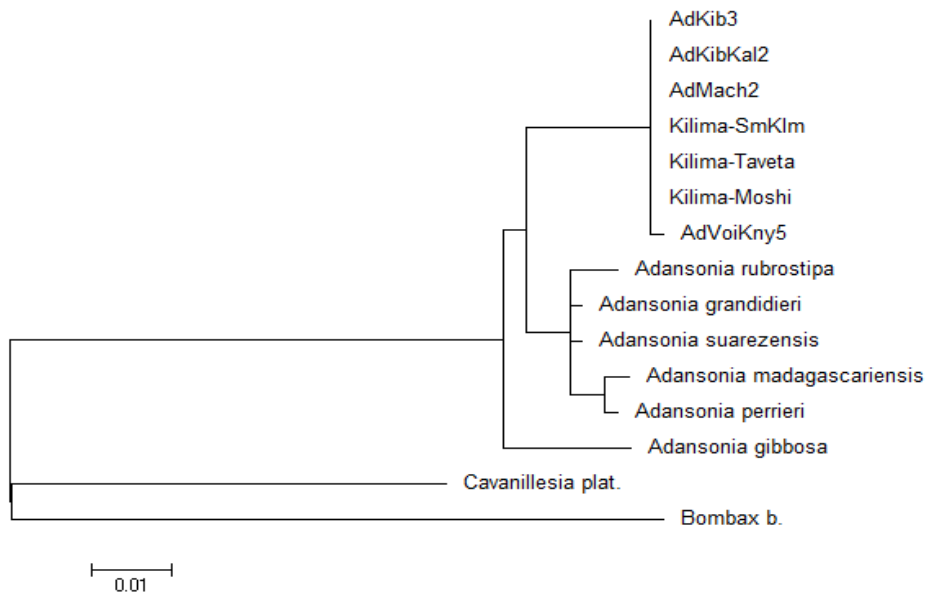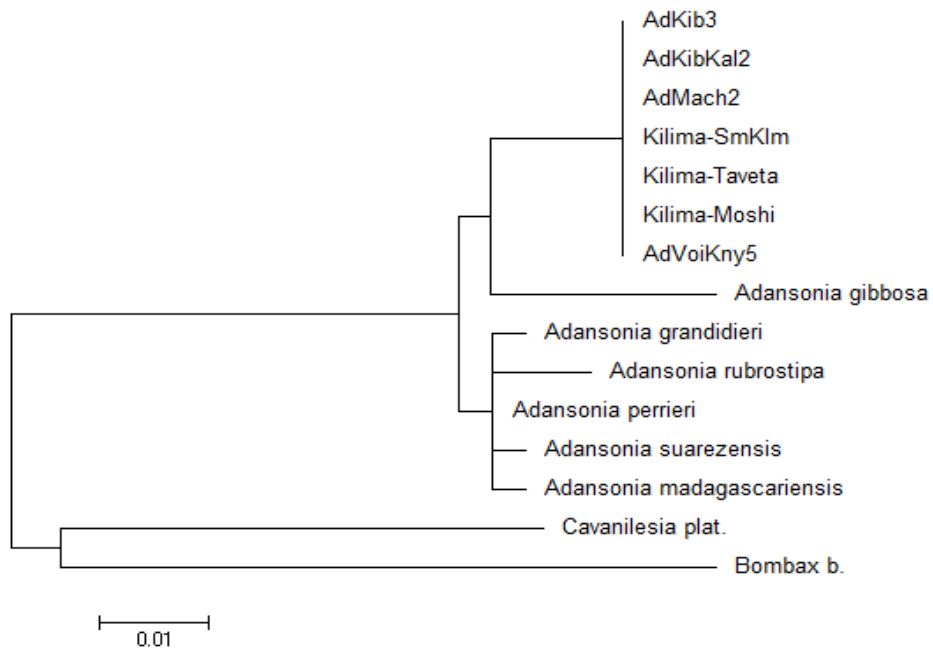 2009). Genetic differentiation among populations is a result of combinations of mutation, selection, migration, mating behavior and genetic drift (Kyndt et al., 2009). Analysis of genetic variation is important for the implementation of breeding and conservation programs and domestication of indigenous tree species (Sreekumar and Renuka, 2006). In the current study on baobab, a total of 335 alleles were amplified from the selected 16 microsatellite loci. The 16 loci were found to be highly polymorphic and displayed up to 4 alleles per sample which is in agreement with the tetraploid nature of *Adansonia digitata.*

The percentage proportion of polymorphic loci is a common genetic diversity parameter in assessing genetic richness. The percentage polymorphic loci amplified in this study ranged from 34.03 to 63.9 % with an average of 50.1 %. (Assogbadjo et al., 2006) reported relatively higher values of percentage polymorphic loci in baobabs studied in West Africa, ranging from 89 to 98 % with a mean of 93 %. High levels of polymorphism suggest high genetic variation within populations (Sreekumar and Renuka, 2006). The results from this study however show low levels of genetic variation within the populations. This could be due to founder effect and human dispersal which could also influence subsequent young populations (Munthali et al., 2012; Pocktsy et al., 2009). Founding baobab populations are suspected to experience genetic drift which lowers genetic diversity in small tree populations (Munthali et al., 2012). Low polymorphism was found in Malindi (34.03%) and Same (34.9%) and may be a sign of genetic erosion. These 2 populations can therefore be considered to be endangered since for a population to survive environmental pressures, it requires a large pool of genetic diversity (Munthali et al., 2012).

Kondo population was found to be the most diverse by having the highest mean diversity (0.194) and this is possibly due to its isolation by distance from the other populations. Populations occurring in varying geographical areas have been found to have different mutation rates (Changadeya et al, 2012) and this could explain the difference in genetic diversity within the 8 populations. Different anthropogenic factors and various natural phenomena e.g. diseases, drought and destruction by wild animals such as elephants have an effect on genetic diversity within populations (Edkins et al., 2007). Differences in tree flowering times within and among populations may also lead to varying degrees of random drift which directly impacts genetic diversity (Dawson et al., 2009). Kondo population, being most diverse, should therefore be given prominence in *in-situ* conservation.

The total genetic differentiation coefficient of the 8 *Adansonia* populations was 0.042 which shows very little genetic differentiation among the populations. This relatively low level of differentiation between the populations corresponds to general observations made for out-crossing plants and long-lived perennials (Nybom & Bartish, 2000). (Bussell, 1999) summarized AMOVA results for RAPD studies on out-breeding vs. selfing plants. He classified plants with higher than 0.45 ϕ_{ST} values as inbreeders and plants with low ϕ_{ST} values of less than 0.413 as outbreeders. A value of 0.042 was observed for the populations in this study confirming baobab to be outbreeding. However, SSRs and not RAPD were used in this study and hence the estimates of variability can be influenced to a certain degree by the marker system utilized (Assogbadjo et al., 2006). Kelly et al., (2004) reported that low population differentiation is caused by extensive gene flow. This may explain the low genetic differentiation among the 8 studied populations which may also be caused by presence of common alleles being shared among the populations (Munthali et al., 2012). Unlike in diploid species, tetraploid species have gene flow events

involving the movement of twice the number of genes (Nassar et al., 2003). This leads to low differentiation among the populations.

The analysis of molecular variance (AMOVA) showed high levels of genetic variation within populations (96%) and very low genetic variation among populations (4%). These results prove the fact that long-lived tree species for example baobab, show low levels of genetic diversity among populations and high levels within populations (Kyndt et al., 2009). The low genetic variation could be due to high levels of gene flow among the populations. This is evidence of a high dispersal ability of the species and abundance of baobab tree pollinators in the region. It is also possible that the studied populations expanded from a recent introduction or from a small isolated population (Bell et al., 2014). Gene flow may occur through dispersal by various vectors for example long distance dispersal of baobab fruits and seeds through mammals mainly humans, bush babies, monkeys and elephants. Recorded baobab pollinators e.g. birds, hawk moths, fruit bats and other insects (Baum, 1995b; Bowman, 1997) may facilitate the gene flow of the paternal line through pollination. The semi-domesticated status of the baobab species could also lead to high levels of gene flow. Traders, rural populations and farmers play a role in gene flow by facilitating village-to-village transport/movement of the baobab fruit between geographically distant populations. Undomesticated populations as a result, have higher levels of within population diversity than domesticated populations (Munthali et al., 2012).

5.2 Population structure and genetic relationships

The Principal Component Analysis (PCoA) demonstrated weak differentiation between the populations with substantial overlap between adjacent clusters in coordinate space. The analysis showed that the genetic structure across the Kenya/Tanzania transect is limited and close similarity between individuals from the Kenya and Tanzania populations. However, individuals belonging to Kondo population exhibited a relatively

45

scattered distribution hence contained the richest genetic information. There was an overlap between Kibwezi, Mtito, Machinery and Voi populations and this suggested a close genetic relationship between the populations. Kondo and Malindi populations also overlapped showing that they have high genetic similarity.

The UPGMA dendogram and Neighbour Joining tree constructed in Darwin 6 were used to analyze the genetic relationships and clustering of the baobab tree individuals. The UPGMA dendogram grouped the tree individuals into two distinct clusters (C1 and C3) and one mixed group (C2). Cluster 1 consisted mainly of Kenyan populations. It contained individuals from Kibwezi, Voi and Mtito Andei. Same-Kilimanjaro individuals also grouped in this cluster. Cluster 2 was the mixed group and it contained individuals from Voi, Kibwezi, Moshi and Machinery. Cluster 3 contained mainly individuals from Malindi and Kondo populations. These 3 clusters were also observed in the Neighbour Joining radial tree where the individuals clustered in a similar way. In support of the findings from the PCA analysis, Kibwezi, Voi and Mtito Andei populations were found to have high genetic similarity. This was expected as they are geographically located relatively close to each other. There is probability of high fruit and seed exchange between farmers and an abundance of baobab tree pollinators leading to a high level of gene flow within the populations as observed by Assogbadjo et al., (2006). Malindi and Kondo populations were also found to be genetically similar. Although these populations are found in different countries, they are geographically close to each other and occur in the same elevation hence the genetic similarity. Mixture of tree individuals from different populations in the same cluster, as in cluster 2, may suggest that the individuals came from the same refugia (origin) or multiple introductions from several refugia. Populations occurring in different geographical locations may not have undergone significant differentiation. The genotypes in the baobab populations could have been mixed through seed dispersal by human influence (Munthali et al., 2012). Similar results were reported by (Assogbadjo et al., 2006) and Pocktsy et al., (2009) where the former reported that

some genotypes of *A. digitata* populations which were growing in different climatic zones in Benin, belonged to more than one gene pool while the latter also found some *A. digitata* haplotypes belonging to more than one region which suggested possibility of human introductions to the areas involved.

The 3 clusters found may be considered as ecotypes of baobabs existing in the Kenya/Tanzania transect. Ecotypes are generally considered to be populations adaptable to specific habitats which could have an implication to conservation, seed distribution, domestication and tree breeding strategies (Assogbadjo et al., 2006). Based on these results some general correlation was observed between clustering and geographical origin and although weakly supported statistically (P value of 0.001) it showed a low degree of genetic structuring in the 8 populations.

5.3 Private alleles

Private alleles are of adaptive importance in the conservation of indigenous trees. The 8 populations studied showed presence of private alleles. Presence of these rare alleles shows that conservation should cover all the populations. Two possible reasons can be used to explain the presence of private alleles. First, mutations may occur *in-situ* and this suggests that the populations being studied have been present for more than one generation. Therefore the existing trees are not founding individuals. Second, the lack of evidence of a recent bottleneck could suggest introductions from various source populations or earlier bottlenecks not detected as deviations from the mutation drift equilibrium. These undetected bottlenecks could be as a result of introduction by ancient humans (Bell et al., 2015).

5.4 Polymorphic Information Content (PIC) Determination

The PIC value is a measure of polymorphism that was introduced by Botstein et al., (1980) to describe a genetic marker's usefulness. It describes the diversity within accessions (intra-population diversity) and characterizes the level of polymorphism in each locus. A PIC value of < 0.25 indicates low polymorphism, a value between 0.25 and 0.5 shows average polymorphism and a value higher than 0.5 shows high polymorphism (Liu et al., 2007). All the 18 loci studied had PIC values exceeding 0.5 and hence are highly polymorphic. The highest PIC value of 0.9831 was recorded by 4 markers: Ad02, Ad06, Ad11 and Ad16. These markers therefore had the greatest genetic diversity of all the markers studied and were highly informative.

5.5 Phylogenetic analysis to confirm presence of *Adansonia kilima*

Phylogenetic analysis of the ITS region was not informative enough to distinguish between *A. digitata* and *A. kilima*. Three samples from *A. kilima* and 4 samples from *A. digitata* were used to construct the Phylogenetic trees using their sequenced ITS regions. The phylogenetic trees showed no distinct difference between the two species. No branching was observed to separate the 2 species as they both clustered under the same branch. All 7 samples had the same length of branches meaning that their sequences are similar and almost no evolutionary time has passed between the trees sampled. No speciation separating the 2 species was observed. All the 7 samples were viewed to be more recent diversions. The lack of resolution in the ITS phylogenies probably is due to a low number of parsimony informative characters. A close relationship among the Malagasy species and the monophyly of *Adansonia gibbosa* was however observed. In addition, the ITS phylogeny strongly supported the existence of 3 major subclades within *Adansonia* which are based on three geographical areas namely Madagascar, Australia and Africa. Studies based on a combination of nuclear and chloroplast DNA and morphological characters by Baum et al., (1998) suggested, albeit weakly, that the African and Australian clades were sisters. The phylogenetic tree obtained from ITS S2F/S3R

clearly supported this conclusion although additional data is needed to resolve the relationships found among the main lineages of *Adansonia* (Pettigrew et al., 2012).

CHAPTER SIX
6.0 CONCLUSIONS AND RECOMMENDATIONS

6.1 CONCLUSIONS

The main findings of this study are suggestive of low genetic differentiation caused by high levels of gene flow and a lack of genetic structure. Kondo, Kibwezi and Moshi populations, having the highest diversity values, are the genetic hotspots in the Kenya/Tanzania transect. Conservation should therefore be focused on these populations. The low diversity observed in Malindi and Same populations is a matter of great concern since baobab is a very important agroforestry species in the East African region. Outbreeding species facing climatic and environmental change are at a great risk by having low genetic diversity. Human factors have been hypothesized to have a major contributing role on genetic erosion experienced in baobab populations.

Anthropogenic factors for example time of flowering play a role on genetic erosion and they should be studied further since the mating system is critical in the genetic structure of trees. The clustering observed showed that genotypes may belong to more than one gene pool and this is due to anthropogenic influence, water dispersal and natural distribution.

Phylogenetic analysis of the ITS region was not informative enough to distinguish between *Adansonia digitata* and *Adansonia kilima*. The phylogenetic trees generated did not show any distinct branching and hence presence of *Adansonia kilima* in the Kenya/Tanzania transect was not confirmed.

6.2 RECOMMENDATIONS

The following recommendations are proposed based on the findings of this study:

1. Conservation and breeding strategies involving genetic studies should be developed for Baobab trees. The results and information generated should be used to establish a database for species conservation, breeding strategies, domestication and general management at national and regional level.

2. There is need for further genetic analyses of Baobab populations in the East African region to further understand baobab dispersal and its effects on genetic variation. Low genetic variation means low adaptability and this possess a risk to species survival and overall evolution. Further studies should be carried out to determine how baobab diversity is influenced by human demographics.

3. It is important that the current genetic diversity in all the populations does not decrease further. Efficient genetic resource management should therefore be established in priority zones where conservation efforts can be focused. Baobab seedlings and samplings should also be developed and distributed to farmers.

4. In order to clearly differentiate between the 2 African baobab species, further studies utilizing better techniques should be carried out. Karyotyping analysis can be carried out to confirm the ploidy levels of the 2 species for example *in situ* hybridisation. The tetraploid *Adansonia digitata* would be expected to have 4 sets of chromosomes and 2 sets would be expected in *Adansonia kilima*. Flow cytometry analysis could also be used to estimate the DNA content of the 2 species and the differences in size used to distinguish the 2 species.

REFERENCES

Adrian, P., Karl, F., Rudolf, W., Dana, G., Li, X., Daniel, A. L., & Clark, S. M. (2007). Radiocarbon dating of a very large African baobab. *Tree Physiology*, *27*, 1569–1574. doi:10.1093/treephys/27.11.1569

Assogbadjo, A. E., Kyndt, T., Sinsin, B., Gheysen, G., & Van Damme, P. (2006). Patterns of genetic and morphometric diversity in baobab (Adansonia digitata) populations across different climatic zones of Benin (West Africa). *Annals of Botany*, *97*(5), 819–30. doi:10.1093/aob/mcl043

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S., & Donoghue, M. J. (1995). The ITS region of Nuclear Ribosomal DNA: A valuable source of evidence on Angiosperm Phylogeny. *Annals of Entomological Society of America*, *82*(2), 247–277. Retrieved from http://www.public.asu.edu/~mfwojci/pdfs/Baldwin etal AMBG1995.pdf

Baldwin, B. (1992). Phylognetic utility of the Internal Transcribed Spacers of Nuclear Ribosomal DNA in Plants: An example from the Compositae.

Baum, D. A. (1995a). *A Systematic Revision of Adansonia (Bombacacea) JSTOR: Annals of the Missouri Botanical Garden, Vol. 82, No. 3 (1995), pp. 440-471*. Retrieved February 17, 2015, from http://www.jstor.org/discover/10.2307/2399893?sid=21105883865243&uid=2&uid=4

Baum, D. A. (1995b). The Comparative Pollination and Floral Biology of Baobabs. *Missouri Botanical Garden*, *82*(2), 322–348.

Baum, D. A., Dewitt, S., Stacey, Y., William, S., Nyffeler, R., Alan, A. L., & Oldham, R. L. (2004). Phylogenetic relationships of Malvatheca (Bombacoideae and Malvoideae; Malvaceae sensu lato) as inferred from plastid DNA sequences. *American Journal of Botany*, *91*(11), 1863–71. doi:10.3732/ajb.91.11.1863

Baum, D. A., & Oginuma, K. (1994). A Review of Chromosome Numbers in Bombacaceae with New Counts for Adansonia. *Taxon*, *43*(1), 11–20.

Baum, D. A., Small, R. L., & Wendel, J. F. (1998). Biogeography a nd Floral Evolution of Baobabs (Adansonia, Bombacaceae) as Inferred From Multiple Data Sets, *47*(2).

Bell, K. L., Rangan, H., Fowler, R., Kull, C. A., Pettigrew, J. D., Vickers, C. E., & Murphy, D. J. (2014). Genetic diversity and biogeography of the boab Adansonia gregorii (Malvaceae: Bombacoideae). *Australian Journal of Botany*, *62*(2), 164– 174. doi:10.1071/BT13209

Bell, K. L., Rangan, H., Kull, C. A., Murphy, D. J., Botanic, R., & Victoria, G. (2015). The history of introduction of the African baobab Malvaceae : Bombacoideae) in the Indian subcontinent Author for correspondence :

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, *32*(3), 314–31. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1686077&tool=pmcent rez&rendertype=abstract

Bowman, D.M.J.S. (1997). Observations on the Demography of the Australian Boab Adansonia gibbosa) in the North-west of the Northern Territory, Australia. *Australian Journal of Botany*, *45*, 893–904.

Bussell, J. (1999). The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of Isotoma petraea (Lobeliaceae). *Molecular Ecology*, *8*(5), 775–789.

Changadeya, W., Kaunda, E., & Ambali, A. J. D. (2012). Molecular characterisation of Musa L. cultivars cultivated in Malawi using microsatellite markers. *African Journal of Biotechnology*, *11*(18), 4140–4157. doi:10.5897/AJB10.1738

Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. M. (2006). Microsatellites and

their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, *255*(1-4), 1–29. doi:10.1016/j.aquaculture.2005.11.031

Cruywagen, E. M., de Beer, Z. W., Roux, J., & Wingfield, M. J. (2010). Three new Graphium species from baobab trees in South Africa and Madagascar. *Persoonia: Molecular Phylogeny and Evolution of Fungi*, *25*, 51–71. doi:10.3767/003158510X550368

Dawson, I. K., Lengkeek, A., Weber, J. C., & Jamnadass, R. (2009). Managing genetic variation in tropical trees: linking knowledge with action in agroforestry ecosystems for improved conservation and enhanced livelihoods. *Biodiversity and Conservation*, *18*(4), 969–986.

De Caluwé, E., Halamová, K., & Van Damme, P. (2009). Baobab (Adansonia digitata L.): A review of traditional uses, phytochemistry and pharmacology. *ACS Symposium Series*, *1021*(1), 51–84. doi:10.1021/bk-2009-1021.ch004

Doyle J.J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue Title. *Phytochemical Bulletin*, *19*, 11–15.

Edkins, M.T., Kruger, L.M., Harris, K., & Midgley, J. J. (2007). Baobabs and elephants in Kruger National Park: nowhere to hide. *African Journal of Ecology*, *46*(2), 119–125.

Edwards, A., Civitello, A., Hammond, H. A., & Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics*, *49*(4), 746–756.

Gebauer, J., El-Siddig, K., & Ebert, G. (2002). Baobab (Adansonia digitata L .): a Review on a Multipurpose Tree with Promising Future in the Sudan. *Gartenbauwissenschaft*, *67*(4), 155–160.

Gebauer, J., & Luedeling, E. (2013). A note on baobab (Adansonia digitata L.) in

Kordofan, Sudan. *Genetic Resources and Crop Evolution*, *60*(4), 1587–1596. doi:10.1007/s10722-013-9964-5

Gebauer, J., & Osman, O. (2004). Fruit production in Sudan, including wild fruits. *Erwerbs-Obstbau*, *46*, 110– 116.

Gupta, P. K., Balyan, H. S., Sharma, P. C., & Ramesh, B. (1996). Microsatellites in plants: A new class of molecular markers. *Current Science*, *70*(1), 45–54.

Gur-arie, R., Cohen, C. J., Eitan, Y., Shelef, L., & Hallerman, E. M. (2000). Simple Sequence Repeats in Escherichia coli : Abundance , Simple Sequence Repeats in Escherichia coli : Abundance , Distribution , Composition , and Polymorphism, 62– 71. doi:10.1101/gr.10.1.62

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium*, (41), 95–98.

Hamada, H., Petrino, M. G., & Kakunaga, T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(21), 6465–6469. doi:10.1073/pnas.79.21.6465

Harris, B. J., & Baker, H. G. (1959). Pollination of flowers by bats in Ghana. *Nigerian Field*, *24:*, 151–159.

Hershkovitz, M. A., & Zimmer, E. A. (1996). Conservation patterns in angiosperm rDNA ITS2 sequences. *Nucleic Acids Research*, *24*(15), 2857–2867. doi:10.1093/nar/24.15.2857

Hillis, D. M., Moritz, C., Porter, C. A. & Baker, R. J. (1991). Evidence for Biased Gene Conversion in Concerted Evolution of Ribosomal DNA. *SCIENCE*, *251*, 308–310.

Hillis, D.M., & Dixon, M. T. (1991). RIBOSOMAL DNA : MOLECULAR EVOLUTION AND PHYLOGENETIC INFERENCE. *The Quarterly Review of Biology*, *66*, 411–453.

Jama, B. A., Mohamed, A. M., Mulatya, J., & Njui, A. N. (2008). Comparing the "'big five'": a framework for the sustainable management of indigenous fruit trees in the drylands of east and central Africa., *Ecol Indic*, 170–179.

Jump, A. S., Marchant, R., & Peñuelas, J. (2009). Environmental change and the option value of genetic diversity. *Trends in Plant Science*, *14*(1), 51–58. doi:10.1016/j.tplants.2008.10.002

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., & Dhawan, A. K. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica*, *177*(3), 309–334. doi:10.1007/s10681-010-0286-9

Kelly, B. A., Hardy, O., & Bouvet, J.M. (2004). Temporal and spatial genetic structure in Vitellaria paradoxa (shea tree) in an agroforestry system in southern Mali. *Molecular Ecology*, *13*(5), 1231–1240.

Kress, J. W., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., & Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(23), 8369–74. doi:10.1073/pnas.0503123102

Kyndt, T., Assogbadjo, A. E., Hardy, O. J., Glele Kakaï, R., Sinsin, B., Van Damme, P., & Gheysen, G. (2009). Spatial genetic structuring of baobab (Adansonia digitata, Malvaceae) in the traditional agroforestry systems of West Africa. *American Journal of Botany*, *96*(5), 950–7. doi:10.3732/ajb.0800266

Larsen, A. S., Vaillant, A., Verhaegen, D., & Kjær, E. D. (2009). Eighteen SSR-primers for tetraploid Adansonia digitata and its relatives. *Conservation Genetics Resources*, *1*(1), 325–328. doi:10.1007/s12686-009-9075-y

Liu, K., & Muse, S. V. (2005). Powermarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, *21*(9), 2128–2129. doi:10.1093/bioinformatics/bti282

Liu, Z. P., Liu, G. S., & Yang, Q. C. (2007). A novel statistical method for assessing SSR variation in autotetraploid alfalfa (Medicago sativa L.). *Genetics and Molecular Biology*, *30*(2), 385–391. doi:10.1590/S1415-47572007000300015

Livingstone, D. (1861). A popular account of missionary travels and researches in South Africa. *London: Murray*.

Mace, G. M., Gittleman, J. L., & Purvis, A. (2003). Preserving the Tree of LIfe. *SCIENCE*, *300*, 1707–1709.

May, R. M. (1994). Conceptual aspects of the quantification of the extent of biological diversity. *Philosophical Transactions of the Royal Society of London*, (345), 13–20.

Miller, M. P. (1997). Tools for Population Genetic Analyses (TFPGA): A Windows® program for the analysis of allozyme and molecular population genetic data, 1–33. doi:10.1111/j.1751-0813.1997.tb15381.x

Munthali, C. R. Y., Chirwa, P. W., Changadeya, W. J., & Akinnifesi, F. K. (2012). Genetic differentiation and diversity of Adansonia digitata L (baobab) in Malawi using microsatellite markers. *Agroforestry Systems*, *87*(1), 117–130. doi:10.1007/s10457-012-9528-2

Nassar, J. M., Hamrick, J. L., & Fleming, T. H. (2003). Population GENETIC STRUCTURE OF VENEZUELAN CACTI (C ACTACEAE) 1. *American Journal of Botany*, *90*(11), 1628–1637. doi:10.3732/ajb90.11.1628

Ndoye, A. L., Diome, T., Gueye, M. C., Sembene, M., & Sy, M. O. (2013). Genetic diversity and demographic evolution of baobab (Adansonia digitata L ., Bombacoideae , Malvaceae) populations in Senegalese Sahelian areas, *12*(38), 5627–5639. doi:10.5897/AJB2013.12431

Nybom, H., & Bartish, I. V. (2000). Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. *Perspectives in Plant Ecology, Evolution and Systematics*, *3*(2), 93–114.

Peakall, R., & Smouse, P. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *MOLECULAR ECOLOGY NOTES*, *6*(1), 288–295.

Pettigrew, J. D., Bell, K. L., Bhagwandin, A., Grinan, E., Jillani, N., Meyer, J., & Vickers, C. E. (2012). Morphology, ploidy and molecular phylogenetics reveal a new diploid species from Africa in the baobab genus Adansonia (Malvaceae: Bombacoideae). *Taxon*, *61*, 1240–1250.

Pocktsy, J. M. L., Lumaret, R., Mayne, D., Vall, A. O. M., Abutaba, Y. I. M., Sagna, M., & Danthu, P. (2009). Chloroplast DNA phylogeography suggests a West African centre of origin for the baobab, Adansonia digitata L. (Bombacoideae, Malvaceae). *Molecular Ecology*, *18*(8), 1707–1715. doi:10.1111/j.1365-294X.2009.04144.x

Prasanth, V.P., & Chandra, S. (1997). ALS-Binary. A program for converting allele size in microsatellite markers into 0-1 (binary) data (ICRISAT, Hyderabad, India).

Prasanth, V. P., Chandra, S., Hoisington, D. A., & Jayashree, B. (1997). Allelobin : a program for allele binning of microsatellite markers . AlleloBin : A Program for Allele Binning in Microsatellite Markers. *International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)*, (January). doi:10.13140/2.1.4397.2165

Provan, J., Powell, W., & Hollingsworth, P. M. (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution*, *16*(3), 142–147. doi:10.1016/S0169-5347(00)02097-8

Raebild, A., Ouedraogo, M., De Groote, S., Van Damme, P., Bayala, J., Sanou, H., & Kjaer, E. D. (2011). Advances in domestication of indigenous fruit trees in the West African Sahel. *New Forest*, (41), 297–315.

Sanchez, A. C. (2011). The baobab tree in Malawi. *Fruits*, *66*(6), 405–416. doi:10.1051/fruits/2011055

Sang, T., Crawford, D. J., & Stuessy, T. F. (1995). Documentation of reticulate

evolution in peonies (Paeonia) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(15), 6813–6817.

Sidibe, M., & Williams, J. T. (2002). Baobab – Adansonia digitata. Fruits for the future. *International Centre for Underutilised Crops, Southampton, UK, 96p.*

Soranzo, N., Provan, J., & Powell, W. (1999). An example of microsatellite length variation in the mitochondrial genome of conifers. *Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada*, *42*(1), 158–161. doi:10.1139/gen-42-1-158

Sreekumar, V. B., & Renuka, C. (2006). Assessment of genetic diversity in Calamus thwaitesii BECC. (Arecaceae) using RAPD markers. *Biochemical Systematics and Ecology*, *34*(5), 397–405.

Start, A. (1972). Pollination of the Baobab (Adansonia digitata) by the fruit bat (Rousethus agyptiacus). *East African Wildlife Journal*, (10).

Taberlet, P., Gielly, L., Pautou, G., & Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, *17*(ii), 1105–1109. doi:10.1007/BF00037152

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. doi:10.1093/molbev/mst197

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, *17*(16), 6463–71. doi:10.1093/nar/17.16.6463

Tautz, D., & Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, *12*(10), 4127–4138.

doi:10.1093/nar/12.10.4127

Teixeira, H., Rodrıguez-Echeverrıa, S., & Nabais, C. (2014). Genetic diversity and differentiation of Juniperus thurifera in Spain and Morocco as determined by SSR. *PLoS ONE*, *9*(2). doi:10.1371/journal.pone.0088996

Tsy, J. M. L. P., Lumaret, R., Flaven-Noguier, E., Sauve, M., Dubois, M. P., & Danthu, P. (2013). Nuclear microsatellite variation in Malagasy baobabs (Adansonia, Bombacoideae, Malvaceae) reveals past hybridization and introgression. *Annals of Botany*, *112*, 1759–1773. doi:10.1093/aob/mct230

Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology*, *23*(1), 48–55. doi:10.1016/j.tibtech.2004.11.005

Wagener, B., Reineke, A., Lohr, B., & Zebitz, C. P. W. (2006). Phylogenetic study of Diadegma species (Hymenoptera : Ichneumonidae) inferred from analysis of mitochondrial and nuclear DNA sequences. *Biological Control*, *37*(2), 131–140. doi:10.1016/j.biocontrol.2006.01.004

Wang, M. L., Barkley, N. A., & Jenkins, T. M. (2009). Microsatellite markers in plants and insects. Part I: applications of biotechnology. *Genes, Genomes And, Genomics*, *3*, 54–67. Retrieved from http://caes.uga.edu/departments/ent/pubs/jenkins_MicrosatelliteI.pdf

Wendel, J. F., Schnabel, A., & Seelanan, T. (1995). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (Gossypium), *92*(January), 280–284.

White, T. J., Bruns, S., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications*. doi:citeulike-article-id:671166

Wiehle, M., Prinz, K., Kehlenbeck, K., Goenster, S., Mohamed, S. A., Finkeldey, R., &

Gebauer, J. (2014). The African baobab (Adansonia digitata, Malvaceae): Genetic resources in neglected populations of the Nuba Mountains, Sudan 1. *American Journal of Botany*, *101*(9), 1498–1507. doi:10.3732/ajb.1400198
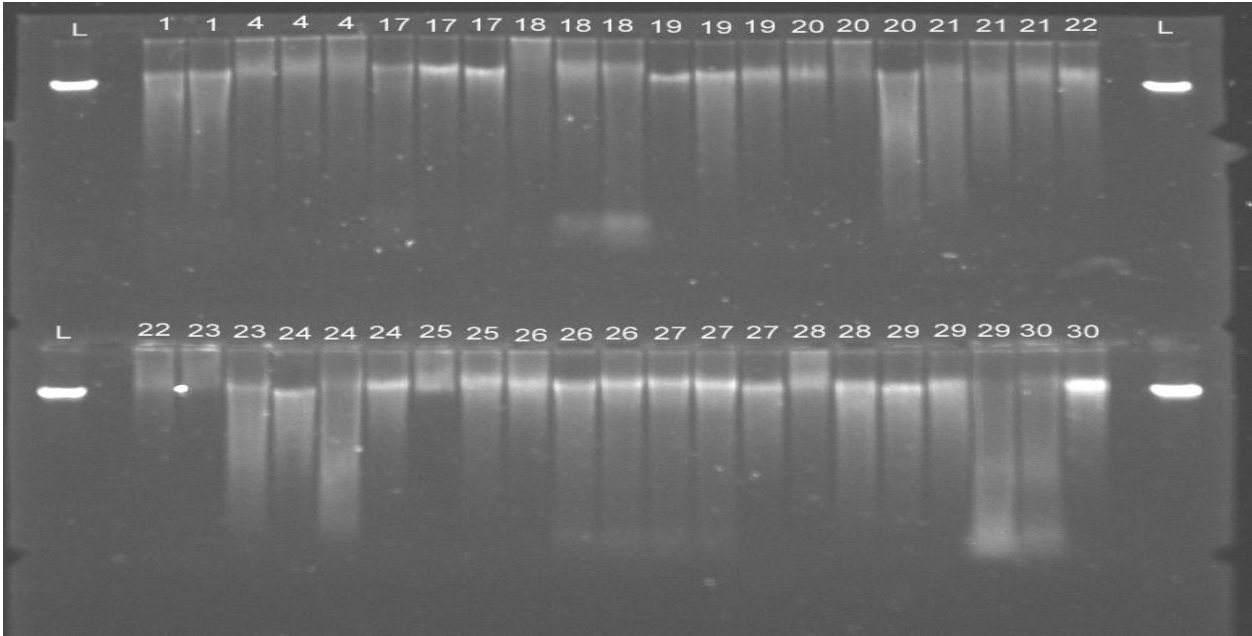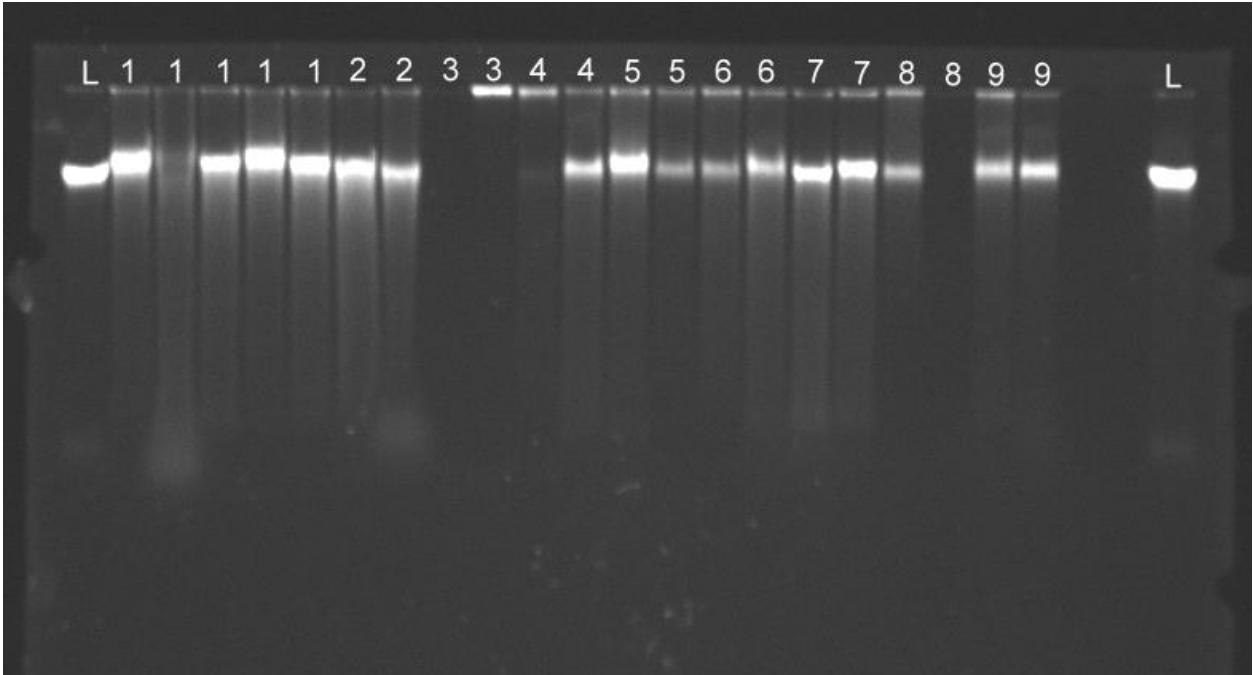
APPENDIX

Appendix 1: Details of baobab provenances sampled for the study.

Tree ID	Location	Altitude	S°	E°
1	VOI KONENYI 003	752	3.35938	38.46979
2	VOI MRARU 004	689	3.3345	38.48879
3	KIBWEZI MANYANGA 001	864	2.38274	37.96258
4	MACHINERY 003	888	2.50631	38.04465
5	MACHINERY 004	873	2.478	38.05601
6	MACHINERY 005	866	2.47662	38.05587
7	KIBWEZI KALUNGU 001	842	2.36643	37.95917
8	KIBW MANYANGA 005	994	2.46394	37.95881
9	KIBW 001	947	2.39389	37.9533
10	MACHINERY 002	899	2.51611	38.03915
11	MACHINERY 001	905	2.51674	38.03981
12	KIBWEZI MANYANGA 003	980	2.46404	37.97276
13	KIBWEZI KALUNGU 002	941	2.36683	37.95912
14	KIBWEZI 003	939	2.38686	37.95565
15	KIBWEZI KALUNGU 004	935	2.35792	37.9579
16	KIBWEZI KALUNGU 005	930	2.35878	37.9764
17	KIBWEZI MANYANGA 004	984	2.46404	37.97277
18	KIBWEZI 005	933	2.38296	37.96239
19	KIBWEZI 004	912	2.3871	37.95606
20	KIBWEZI 002	955	2.39036	37.9534
21	KIBWEZI KALUNGU 003	963	2.35761	37.9583
22	KIBWEZI MANYANGA 002	932	2.43496	37.96457
23	MTITO ANDEI 002	775	2.67906	38.14894
24	KATHEKANI MTITO 002	753	2.65904	38.1413

25	MTITO ANDEI 004	817	2.68157	38.13905
26	KATHEKANI MTITO 005	791	2.65669	38.14848
27	MTITO ANDEI 001	766	2.67903	38.14893
28	KATHEKANI MTITO 003	806	2.65606	38.1451
29	KATHEKANI MTITO 001	754	2.6752	38.13842
30	MTITO ANDEI 005	816	2.68306	38.13941
31	MTITO ANDEI 003	800	2.68335	38.14243
32	KATHEKANI MTITO 006	781	2.65192	38.15166
33	VOI KONENYI 001	743	3.36476	38.46865
34	VOI MRARU 005	689	3.34336	38.48878
35	VOI KONENYI 002	747	3.36332	38.46881
36	VOI KONENYI 004	758	3.37237	38.4743
37	VOI MRARU 002	675	3.3497	38.49632
38	VOI KONENYI 005	751	3.38574	38.38574
39	VOI MRARU 001	643	3.371	38.50804
40	KONDO BAHARI BICHI		6.38774	39.11939
41	KONDO BAHARI BICHI		6.38743	39.11932
42	KONDO BAHARI BICHI		6.38842	39.1207
43	KONDO BAHARI BICHI-TENGETA		3.57869	37.40686
44	UNUNIO		6.38129	37.20673
45	KEREGE KIWETE-BAGAMOYO		6.34433	39.01573
46	KEREGE KIWETE		6.34209	39.0163
47	KOALE-BAGAMOYO		6.27512	38.56196
48	GAMA- KILIMANJARO	534	4.60759	37.9969
49	SAME-KILIMANJARO	880	4.05404	37.45451
50	SAME	896	4.05237	37.71453
51	SAME	938	4.01983	37.42204

52	MWANGA	978	3.45336	37.36176
53	GHONA CHEKERENI	743	3.44868	37.52927
54	MOSHI	829	3.22816	37.25973
55	MOSHI	827	3.22804	37.25938
56	MOSHI		3.2112	37.20673
57	TAVETA	809	3.36074	37.70629
58	TAVETA	839	3.21139	37.42438
59	MALINDI MANDUGUNI			
60	MALINDI MANDUGUNI			
61	MANDUGUNI JUNCTION			
62	MALINDI MANDUGUNI			

Appendix 2: Gel images of DNA extracted from the leaves sampled.

Appendix 3: Gel image of SSR amplification using Primer Ad01
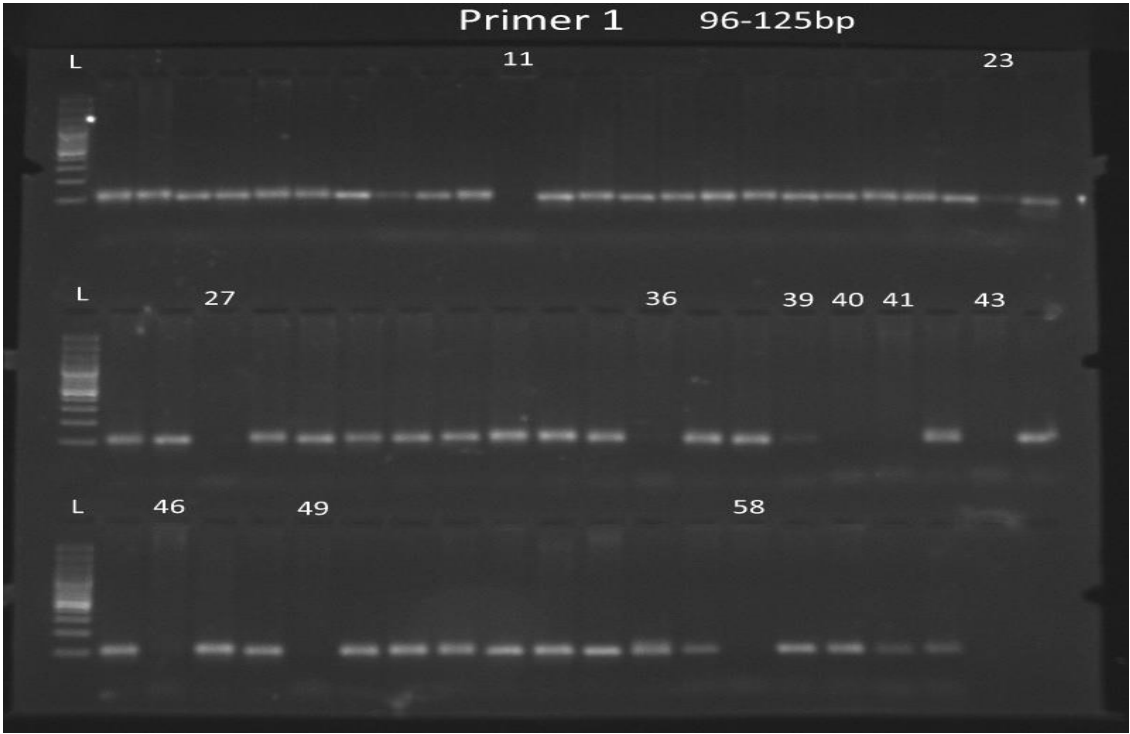
Appendix 4: Gel image of SSR amplification using Primer Ad04
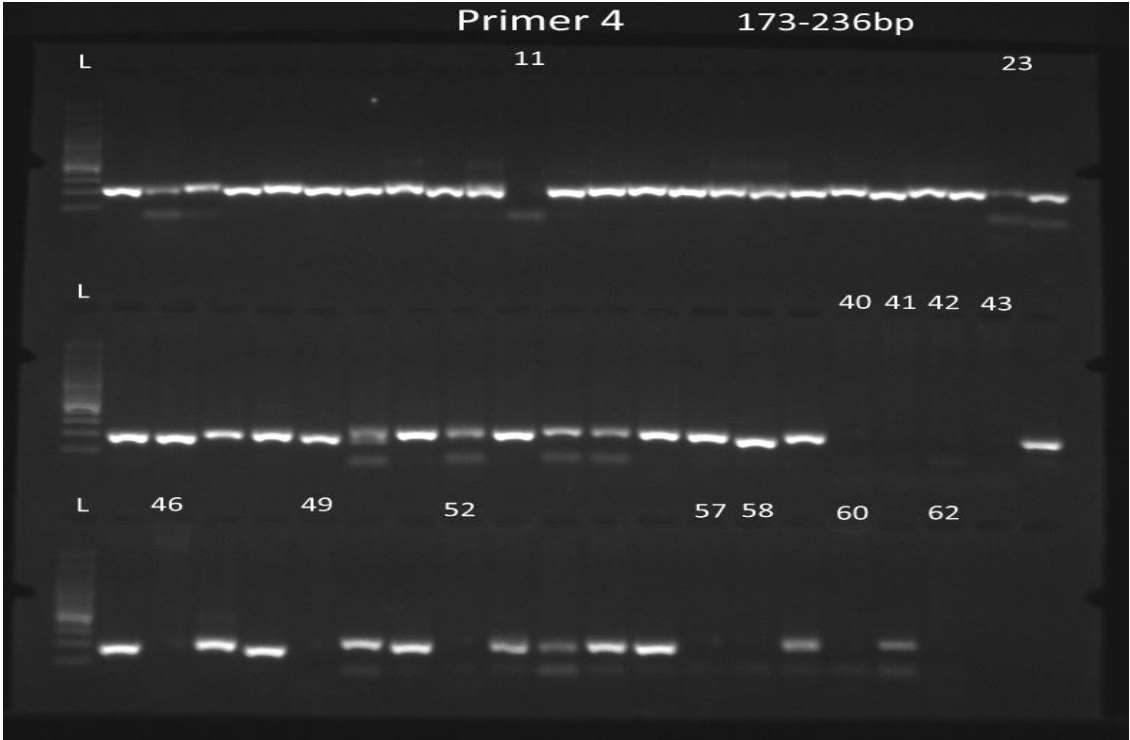
Appendix 5: Gel image of SSR amplification using Primer Ad06
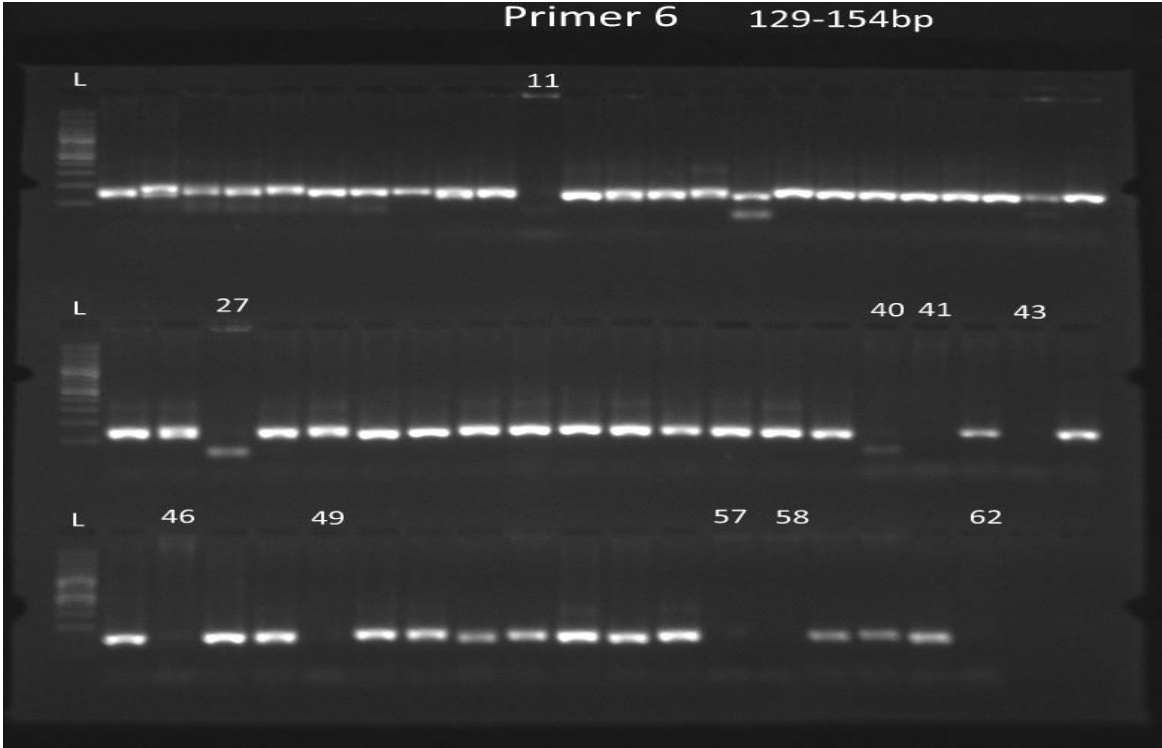
Appendix 6: Gel image of SSR amplification using Primer Ad07
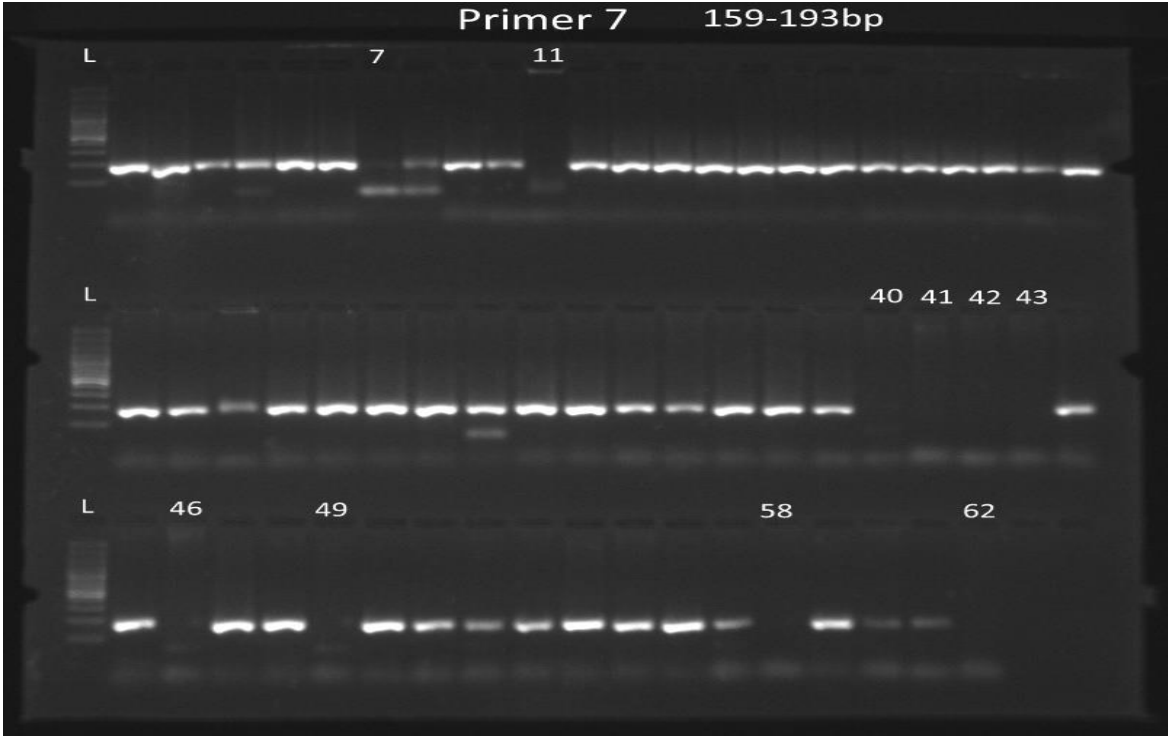
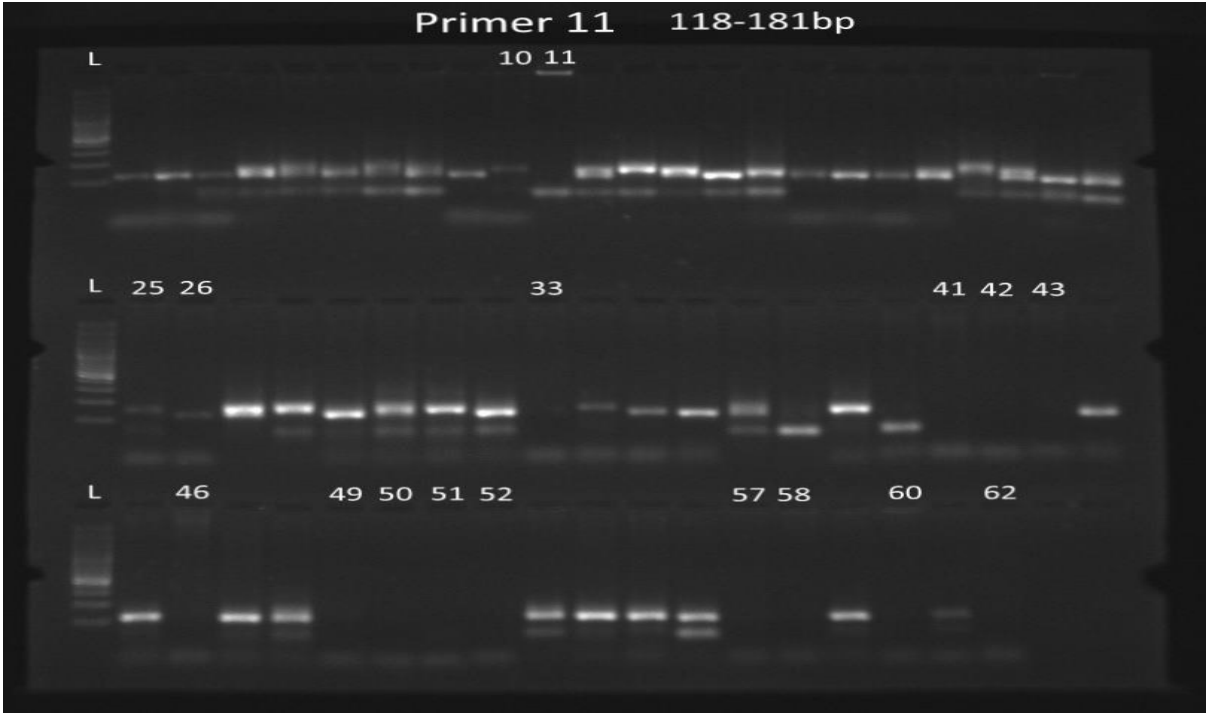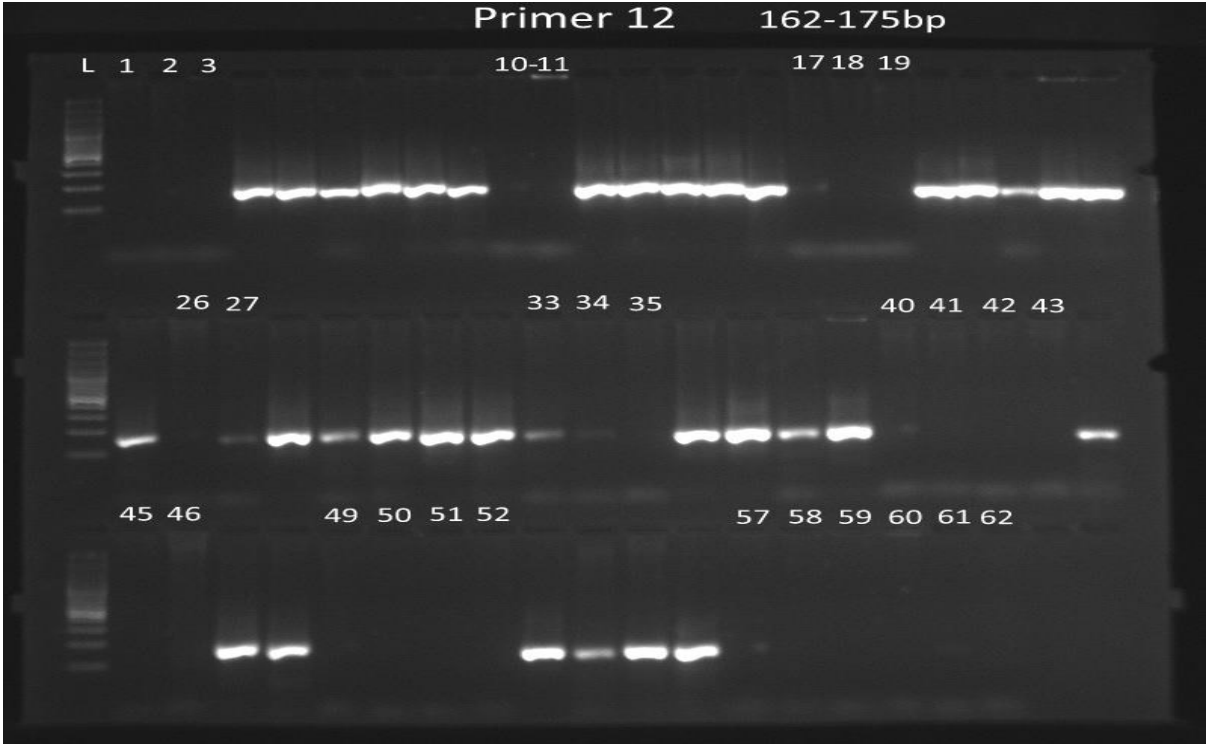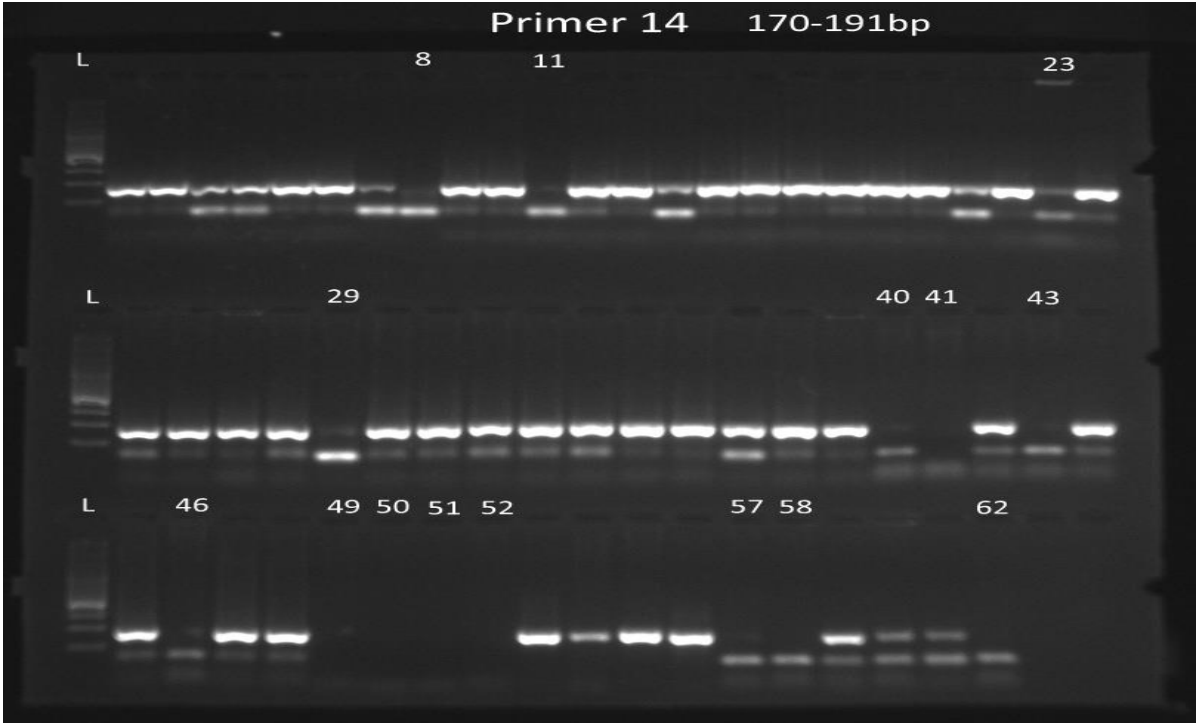**Appendix 7: Gel image of SSR amplification using Primer Ad08**
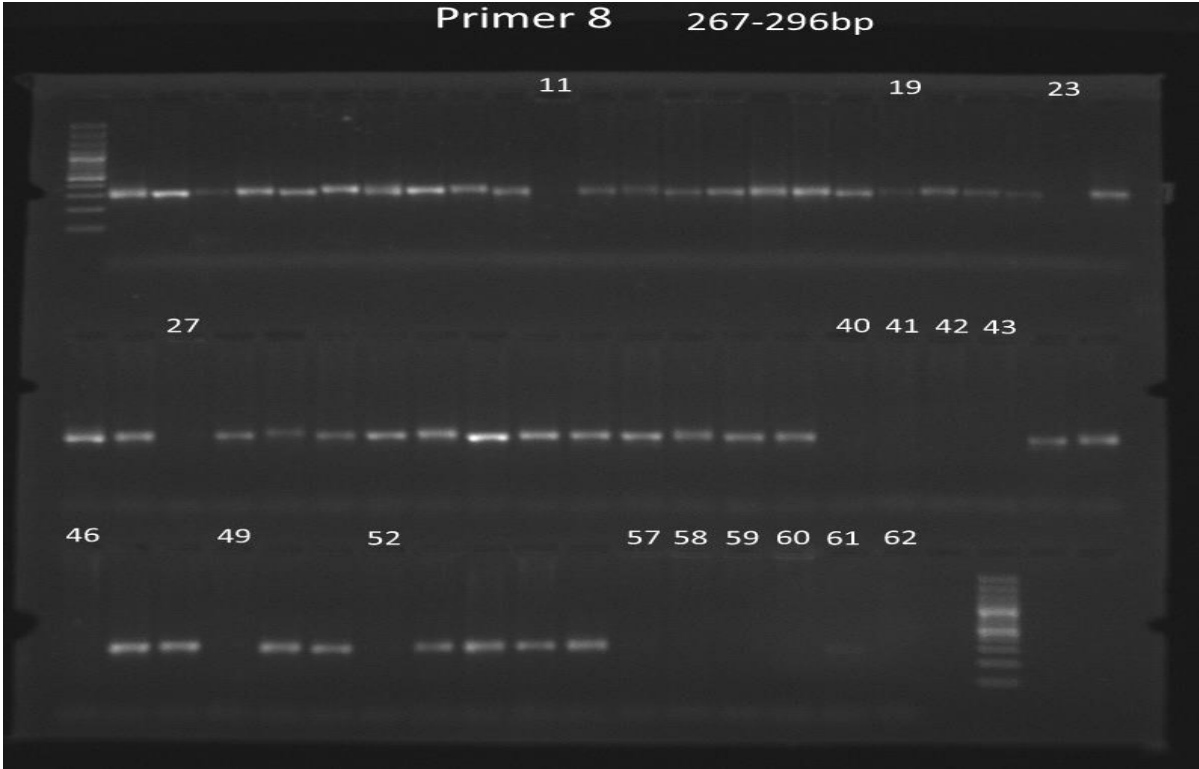
Appendix 8: Gel image of SSR amplification using Primer Ad09
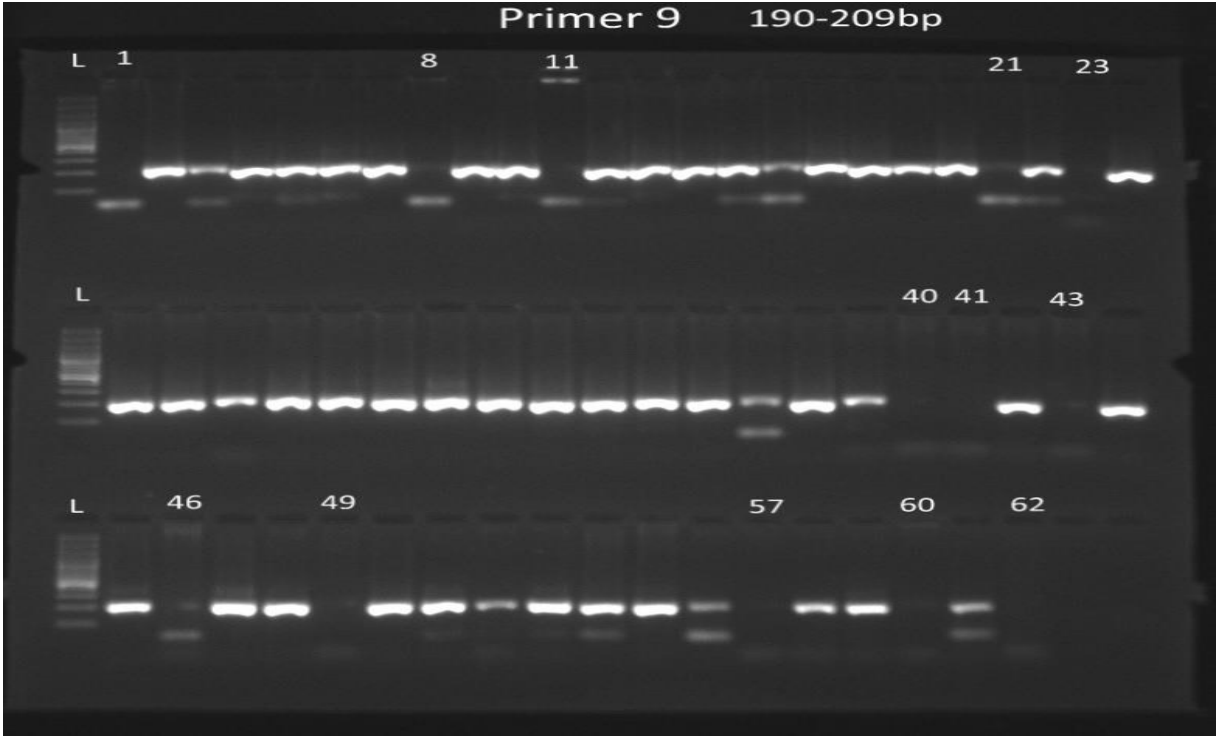
Appendix 9: Gel image of SSR amplification using Primer Ad10
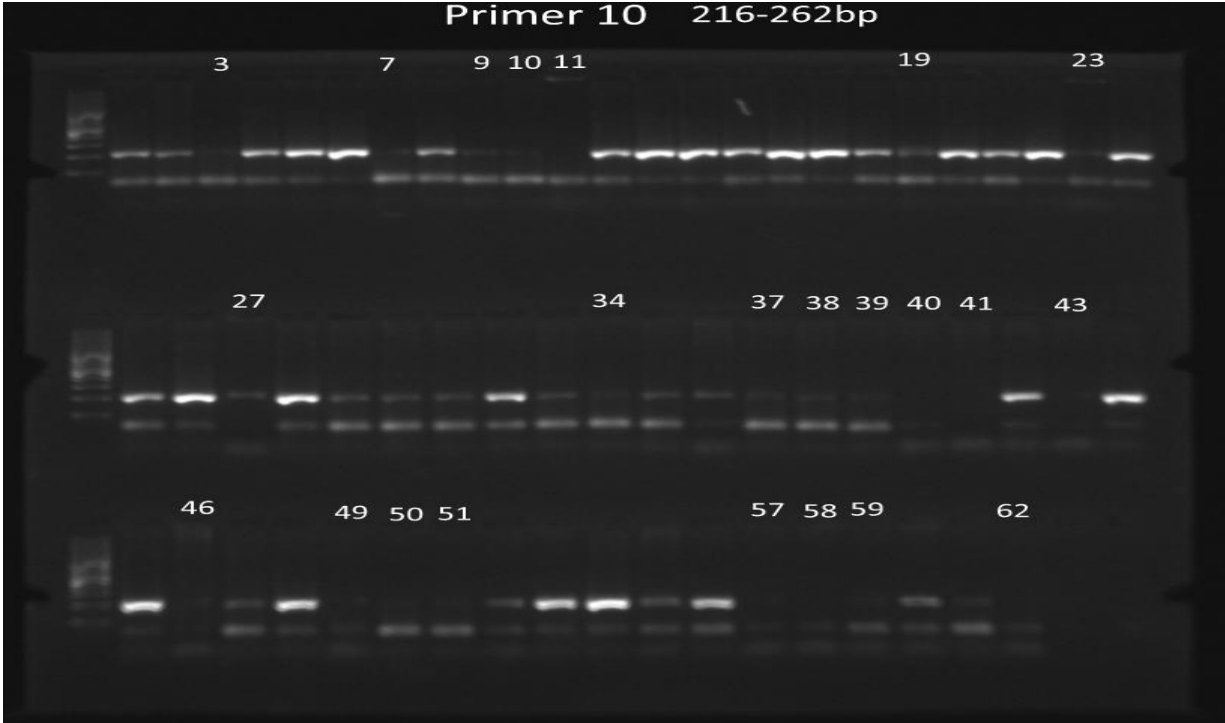
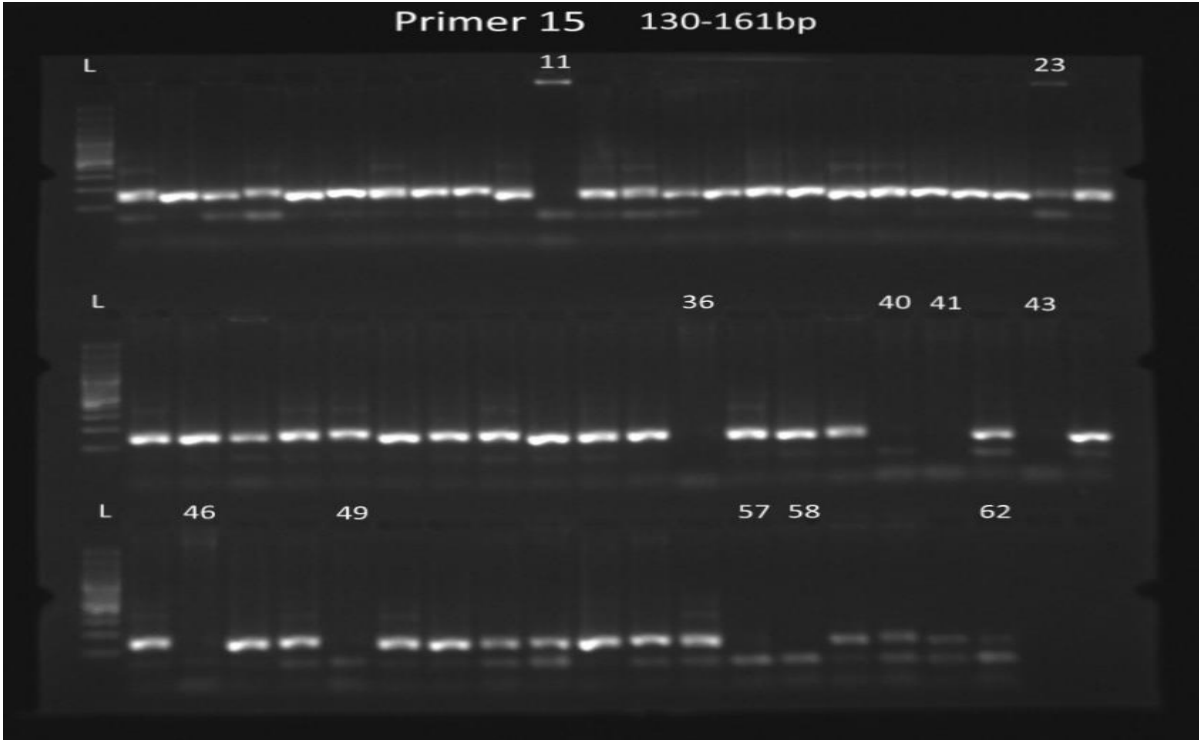**Appendix 10: Gel image of SSR amplification using Primer Ad11**

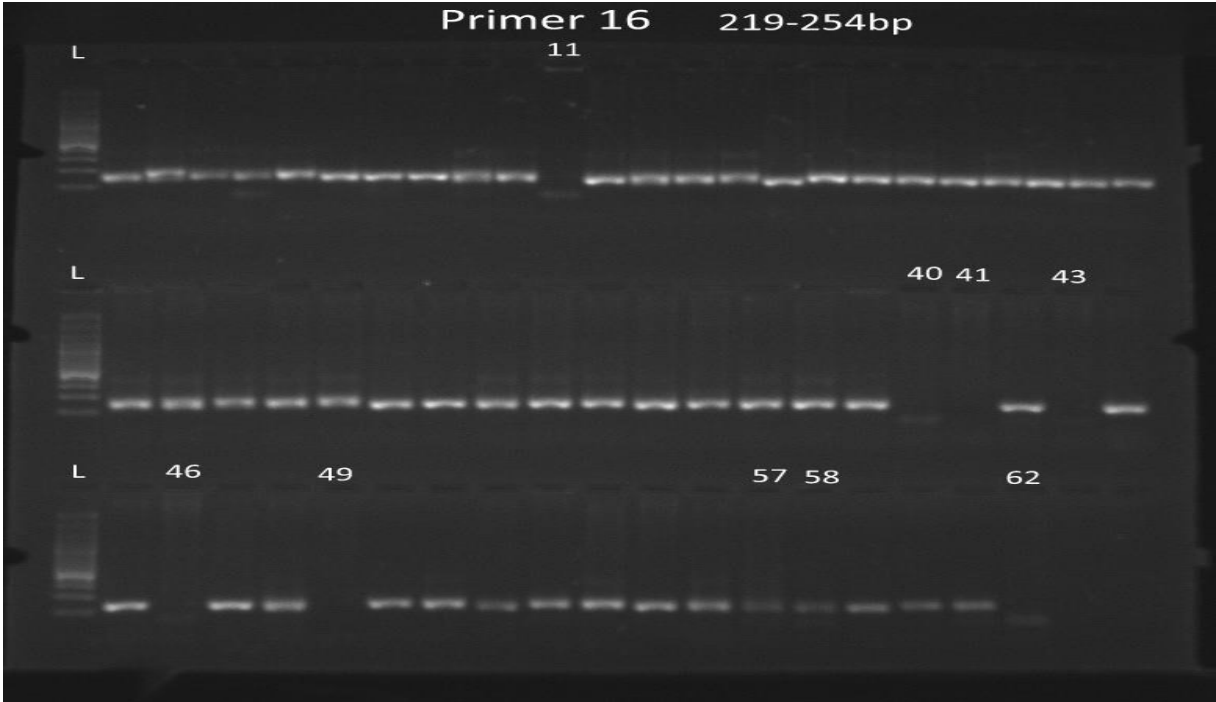**Appendix 11: Gel image of SSR amplification using Primer Ad12**

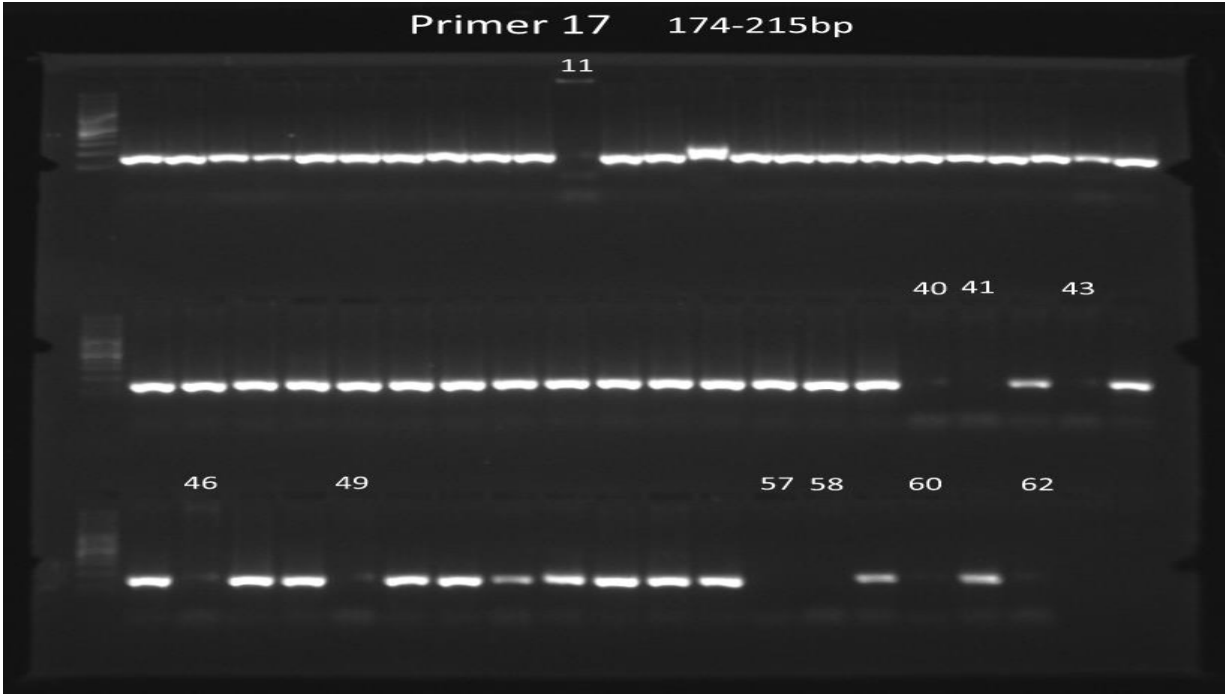**Appendix 12: Gel image of SSR amplification using Primer Ad14**

Appendix 13: Gel image of SSR amplification using Primer Ad15

Appendix 14: Gel image of SSR amplification using Primer Ad16

Appendix 15: Gel image of SSR amplification using Primer Ad17

Appendix 16: Gel image of SSR amplification using Primer Ad18

Appendix 17: Multiple sequence alignment of consensus sequences using the MUSCLE plugin.